Fast Generation of Sound Zones Using Variable Span Trade-Off Filters in the DFT-Domain

Lee, Taewoong; Shi, Liming; Nielsen, Jesper Kjær; Christensen, Mads Græsbøll

Published in:
IEEE/ACM Transactions on Audio, Speech, and Language Processing

DOI (link to publication from Publisher):
10.1109/TASLP.2020.3042701

Publication date:
2021

Document Version
Accepted author manuscript, peer reviewed version

Fast Generation of Sound Zones Using Variable Span Trade-Off Filters in the DFT-Domain

Taewoong Lee, *Student Member, IEEE,* Liming Shi, *Student Member, IEEE,* Jesper Kjær Nielsen, *Member, IEEE* , and Mads Græsbøll Christensen, *Senior Member, IEEE*

***Abstract*—The creation of sound zones with frequency-domain variable span trade-off filters (VAST) is investigated herein. Both narrowband and broadband discrete Fourier transform (DFT)-domain VAST approaches are proposed, and we discuss their relationship to the existing time-domain VAST approach. The core idea in VAST is to apply a generalized eigenvalue decomposition to the spatial statistics to control the trade-off between acoustic contrast and signal distortion. Moreover, a method for determining the optimal Lagrange multiplier that controls this trade-off is also considered in terms of physical, meaningful parameters. Through analysis and experiments, a performance comparison using measured room impulse responses is conducted not only between the two proposed methods but also between the two proposed methods and the existing time-domain approach. The results confirm that the broadband approach is able to transfer the acoustic contrast from one frequency bin to another, which is not the case for the narrowband approach. Furthermore, the results also show that the proposed DFT-domain VAST approach can be considered to be a special case of the time-domain VAST approach.**

***Index Terms*—Optimization, personal sound, sound zones, variable span trade-off filters.**

I. Introduction

In an acoustically shared space, the purpose of a personal sound system is to provide acoustically and spatially separated regions, also known as sound zones, for different audio contents, without the use of headphones [1]–[3]. Alternatively, as originally suggested more than two decades ago [1], sound zones can be created by controlling a set of loudspeakers. Various applications of sound zones have been investigated, e.g, in aircraft [4] and/or car cabins [4]–[7], at outdoor concerts [8], and for mobile phones [9], [10].

The idea behind the creation of sound zones is to control multiple loudspeakers according to different strategies. Typically, a bright zone (or a listening zone) where the desired audio content is being reproduced and a dark zone (or a quiet zone), the acoustic potential energy of which is minimized as much as possible, are considered at the same time. The zones can then be created by maximizing the acoustic potential energy ratio between the bright and dark zones, which is referred to as acoustic contrast, using the acoustic contrast control (ACC) method in [11] that is based on the generalized eigenvalue problem. Alternatively, the zones can be created by minimizing the difference between the desired and reproduced sound fields of the bright and dark zones, which is referred to as the reproduction error, using the pressure matching (PM) method in [12], [13] that is based on the least squares problem. Finally, the zones can also be created by minimizing the reproduction error represented by spatial harmonic expansion, based on the mode matching method in [14]–[16].

ACC is one of the most-popular and best-known methods; however, this approach has multiple drawbacks. First, because ACC does not take the phase information into account, the spatial distribution over the bright zone is difficult to control, which results in significant signal distortion in the bright zone. This issue was later resolved using planarity control [17]. Second, the solution of ACC is the eigenvector corresponding to the largest eigenvalue, such that an additional procedure to find the magnitude of the solution is required, as proposed in [18]–[21]. Third, ACC computes a matrix inversion that is required to solve the generalized eigenvalue problem; this computation can often be highly ill-conditioned, depending on the geometry of the loudspeaker array, the locations of the sound zones, and the frequency of interest. Either to improve the robustness or, in particular, to avoid the ill-conditioned inverse problem, a method that maximizes the energy difference between the bright and dark zones [22], an extensive study on regularization [23], and a geometry optimization problem [24] were investigated. Moreover, ACC was initially proposed in the frequency domain, which solves the problem one frequency at a time, although recently, the time-domain version, the so-called broadband ACC (BACC), was also proposed [25]. However, BACC results in significant signal distortion because only the few frequencies for which a large contrast can be obtained are reproduced. To mitigate this issue, additional constraints were introduced to obtain a uniform frequency response, for example, in terms of the acoustic contrast in [26]–[28]. Furthermore, the time-domain approach typically suffers from a high degree of computational complexity. Note that methods from other domains have also been considered as possible solutions to this challenge, e.g., [5], [29], [30].

In contrast, PM controls both the magnitude and phase information because the desired sound field is defined. Unfortunately, however, this degree of control results in a degradation of the acoustic contrast. To improve the acoustic contrast while preserving the availability to control the phase, the following two methods were proposed. In [31], a hybrid method was proposed that combined PM [13] and the method in [22]. This method provides a trade-off between the acoustic contrast and the degree of phase control. In [18], [32], a method that tunes the importance between the amount of acoustic potential

T. Lee, L. Shi, J. K. Nielsen, and M. G. Christensen are with the Audio Analysis Lab, CREATE, Aalborg University, 9000 Aalborg, Denmark (e-mail: tlee, ls, jkn, mgc@create.aau.dk)

energy in the dark zone and that of the signal distortion in the bright zone was proposed.

Recently, a general and exact framework in the time-domain, which is a subspace-based approach inspired by [33]–[38] in speech enhancement, was proposed [39]. This framework, which is referred to as the variable span trade-off (VAST) filter, allows one to control the trade-off between the acoustic contrast and the signal distortion by easily tuning two user parameters. Because the reproduction error can be moved from one frequency to another frequency in the time-domain approach, perceptually optimized approaches also were studied using the VAST framework in [40], [41]. Here, however, there are several drawbacks. First, because VAST exploits the generalized eigenvalue decomposition in the time-domain, the method suffers from a high degree of computational complexity. This drawback was improved by using the conjugate gradient method in [42]; however, it still requires a fairly large number of computations. Second, the two user parameters were assumed to be chosen by the user; as a result, the performance might not be optimal to the constraint, i.e., the residual energy in the dark zone.

In this paper, two methods based on the VAST framework, although residing in the DFT-domain, are proposed to reduce the computational complexity from which the time-domain approach suffers while preserving the trade-off property of the VAST framework. A thorough comparison between the methods in the DFT-domain and in the time-domain is carried out. This type of comparison was also previously undertaken in [43], [44], although those authors reached different conclusions. The methods from both domains yielded exactly the same results in [44], whereas [43] observed different results. In this paper, we investigate this inconsistency and determine the underlying conditions that explain such results. Lastly, because the constraint on the residual energy in the dark zone is neither intuitive nor physically meaningful, different constraints will be considered, as in [45].

This paper is organized as follows: The sound field modelling, mean squared error (MSE) criterion, and performance evaluation metrics are described in Sec. II. The proposed methods are derived in Sec. III by considering a narrowband approach first, followed by a broadband approach. In Sec. IV, the proposed methods are thoroughly analyzed and discussed. The proposed methods are evaluated and compared to the time-domain approach in Sec. V. The paper concludes in Sec. VI.

II. Problem Statement

A. Notation

Throughout this paper, the following notation is used. Scalar variables, integer constants, vectors, and matrices are denoted by lower case symbols a, upper case symbols A, lower case bold $\boldsymbol{a}$, and upper case bold $\boldsymbol{A}$, respectively. Subscripts $(\cdot)_{\mathrm{B}}$ and $(\cdot)_{\mathrm{D}}$ are used to denote the bright and dark zones, respectively. A subscript $(\cdot)_{\mathrm{C}}$ is a variable that can be either $_{\mathrm{B}}$ or $_{\mathrm{D}}$. When the bright and dark zones are combined, subscript $(\cdot)_{\mathrm{F}}$ is used. In addition, $\underline{(\cdot)}$ represents a vector or a matrix being concatenated, and $\hat{(\cdot)}$ denotes the time-domain quantity. Indices n and k denote the discrete time sample index and the discrete frequency bin index, respectively. The symbol $\otimes$ denotes the Kronecker product operator. $(\cdot)^T$ and $(\cdot)^H$ denote the transpose and conjugate transpose (or Hermitian) operators, respectively. $\mathrm{diag}(\cdot)$ and $\mathrm{blkdiag}(\cdot)$ denote a diagonal matrix and a block diagonal matrix, respectively. The real and complex coordinate spaces with $M \times N$ dimensions are denoted by $\mathbb{R}^{M\times N}$ and $\mathbb{C}^{M\times N}$, respectively.

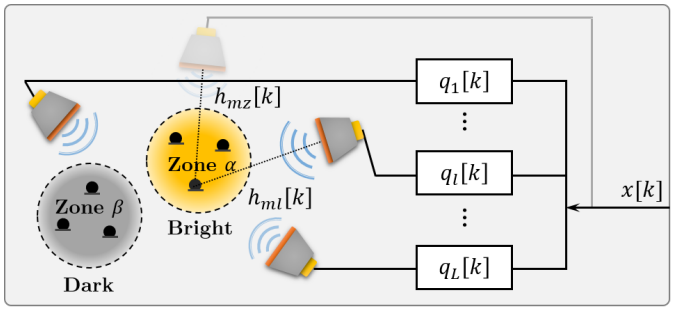


Fig. 1. An illustration of a system geometry of sound zones. The input signal $x[k]$ is fed into L loudspeakers after filtering by the corresponding control filter $q_l[k]$. The transfer function from loudspeaker l to control point m is represented as $h_{ml}[k]$. A bright zone and a dark zone are spatially sampled by M_B and M_D control points, respectively. Virtual source z, depicted in a blurry manner, is introduced, and the corresponding impulse response is represented as $h_{mz}[k]$.

B. Sound field modelling

In this paper, we tackle the problem of creating both a bright zone and a dark zone in an enclosed space, as shown in Fig. 1. Note that a solution for this problem that accounts for each audio content can be exploited to have multiple bright zones via the superposition principle. First, each of the zones is discretized and spatially sampled by M control points, and an input signal $\hat{x}[n]$ of length N that we want to generate at the bright zone is fed into L loudspeakers after being filtered by finite impulse response (FIR) filters for each of L loudspeakers. The filtered signals are emitted through the space, eventually being captured at each of M control points.

In the frequency-domain[1], which is a common approach in sound zone control, e.g., [6], [11], [13], [17], [18], [22], [31], the reproduced sound pressure $p_m[k]$ at control point m can be represented by the transfer function $h_{ml}[k]$ from loudspeaker l to control point m, the input signal $x[k]$, and the control filter $q_l[k]$ such that

$$p_m[k] = x[k] \sum_{l=1}^{L} h_{ml}[k] q_l[k] \tag{1}$$

$$= x[k] \boldsymbol{h}_m^T[k] \boldsymbol{q}[k] \ , \tag{2}$$

where

$$\boldsymbol{h}_m[k] = \left[h_{m1}[k], \ \cdots, \ h_{mL}[k]\right]^T \in \mathbb{C}^{L\times 1} \ , \tag{3}$$

$$\boldsymbol{q}[k] \ = \left[q_1[k], \ \cdots, \ q_L[k]\right]^T \in \mathbb{C}^{L\times 1} \ , \tag{4}$$

[1]It should be noted that the time-domain approach and the frequency-domain approach are only equivalent when a) the input signal is periodic in N, i.e., $\hat{x}[n] = \hat{x}[n + cN]$ where c is an integer, or is infinitely long and b) the length of the control filter is the same as the DFT size, which is N, and it is equal to or longer than the length of the room impulse response. We show a case in which both approaches are equivalent in Sec. V-F1.

Note that the transfer function $h_{ml}[k]$ can either be modelled or measured in advance.

Let us define the total sound field using the subscript $_{\mathrm{F}}$, which is

$$\begin{aligned}\boldsymbol{p}_{\mathrm{F}}[k] &= \begin{bmatrix}\boldsymbol{p}_{\mathrm{B}}^T[k] & \boldsymbol{p}_{\mathrm{D}}^T[k]\end{bmatrix}^T \in \mathbb{C}^{(M_{\mathrm{B}}+M_{\mathrm{D}})\times 1} && (5)\\ &= x[k]\boldsymbol{H}_{\mathrm{F}}[k]\boldsymbol{q}[k]\ , && (6)\end{aligned}$$

where

$$\begin{aligned}\boldsymbol{H}_{\mathrm{F}}[k] &= \begin{bmatrix}\boldsymbol{H}_{\mathrm{B}}^T[k] & \boldsymbol{H}_{\mathrm{D}}^T[k]\end{bmatrix}^T \in \mathbb{C}^{(M_{\mathrm{B}}+M_{\mathrm{D}})\times L}\ , && (7)\\ \boldsymbol{p}_{\mathrm{C}}[k] &= \begin{bmatrix}p_1[k], & \cdots, & p_{M_{\mathrm{C}}}[k]\end{bmatrix}^T && (8)\\ &= x[k]\boldsymbol{H}_{\mathrm{C}}[k]\boldsymbol{q}[k] \in \mathbb{C}^{M_{\mathrm{C}}\times 1}\ , && (9)\\ \boldsymbol{H}_{\mathrm{C}}[k] &= \begin{bmatrix}\boldsymbol{h}_m[k], & \cdots, & \boldsymbol{h}_{M_{\mathrm{C}}}[k]\end{bmatrix}^T \in \mathbb{C}^{M_{\mathrm{C}}\times L}\ . && (10)\end{aligned}$$

where $\boldsymbol{H}_{\mathrm{C}}[k]$ is the transfer function matrix. The acoustic potential energy in the total field can then be modified from (6) as

$$\begin{aligned}e_{\mathrm{F}}[k] &= \boldsymbol{p}_{\mathrm{F}}^H[k]\boldsymbol{p}_{\mathrm{F}}[k] && (11)\\ &= |x[k]|^2\boldsymbol{q}^H[k]\Big(\boldsymbol{H}_{\mathrm{B}}^H[k]\boldsymbol{H}_{\mathrm{B}}[k]+\boldsymbol{H}_{\mathrm{D}}^H[k]\boldsymbol{H}_{\mathrm{D}}[k]\Big)\boldsymbol{q}[k] && (12)\\ &= |x[k]|^2\boldsymbol{q}^H[k]\Big(\boldsymbol{R}_{\mathrm{B}}[k]+\boldsymbol{R}_{\mathrm{D}}[k]\Big)\boldsymbol{q}[k]\ , && (13)\end{aligned}$$

where

$$\boldsymbol{R}_{\mathrm{C}}[k] = \boldsymbol{H}_{\mathrm{C}}^H[k]\boldsymbol{H}_{\mathrm{C}}[k] \quad \text{for C} \in \{\mathrm{B}, \mathrm{D}\}\ , \tag{14}$$

where $\boldsymbol{R}_{\mathrm{B}}[k]$ and $\boldsymbol{R}_{\mathrm{D}}[k]$ are the spatial correlation matrices for the bright and dark zones, respectively. Later, we return to these matrices to jointly diagonalize them by using their properties.

Let us define the desired sound field for the bright and dark zones as

$$\begin{aligned}\boldsymbol{d}_{\mathrm{B}}[k] &= \begin{bmatrix}d_1[k], & \cdots, & d_{M_{\mathrm{B}}}[k]\end{bmatrix}^T\\ &= x[k]\boldsymbol{h}_z[k] \in \mathbb{C}^{M_{\mathrm{B}}\times 1}, && (15)\\ d_m[k] &= x[k]h_{mz}[k], && (16)\\ \boldsymbol{h}_z[k] &= \begin{bmatrix}h_{1z}[k], & \cdots, & h_{M_{\mathrm{B}}z}[k]\end{bmatrix}^T \in \mathbb{C}^{M_{\mathrm{B}}\times 1}, && (17)\\ \boldsymbol{d}_{\mathrm{D}} &= \mathbf{0}_{M_{\mathrm{D}}} \in \mathbb{C}^{M_{\mathrm{D}}\times 1}, && (18)\\ \boldsymbol{d}_{\mathrm{F}}[k] &= \begin{bmatrix}\boldsymbol{d}_{\mathrm{B}}^T[k] & \boldsymbol{d}_{\mathrm{D}}^T[k]\end{bmatrix}^T \in \mathbb{C}^{(M_{\mathrm{B}}+M_{\mathrm{D}})\times 1}, && (19)\end{aligned}$$

where $\mathbf{0}_{M_{\mathrm{D}}}$ is an all-zeros vector of length M_{D}, $\boldsymbol{d}_{\mathrm{B}}[k]$ is the desired sound field for the bright zone, and $h_{mz}[k]$ is the transfer function from the virtual source z to control point m. It should be noted that the position of the virtual source is a design choice affecting the desired sound field in the bright and dark zones. Typically, $\boldsymbol{h}_z[k]$ is defined (or designed) as a plane or a spherical wave field in an anechoic environment emitted by the virtual source [14], [18], [31], which implicitly makes the sound zone control method perform dereverberation to minimize the difference between the desired and reproduced sound fields if a reverberant environment is given [46]–[48][2]. Our task is to design the control filter $\boldsymbol{q}[k]$ to match $\boldsymbol{p}_{\mathrm{F}}[k]$ to $\boldsymbol{d}_{\mathrm{F}}[k]$ as well as possible in a mathematically tractable way.

[2]The inverse filtering technique was considered to equalize the room influence, for example, [46]–[48].

C. Mean squared error (MSE) criteria

Now let us introduce two MSE criteria that measure the distance between the desired and reproduced sound fields for the bright and dark zones, respectively, i.e.,

$$\begin{aligned}\mathcal{S}_{\mathrm{B}}(\boldsymbol{q}[k]) &= ||\boldsymbol{d}_{\mathrm{B}}[k]-\boldsymbol{p}_{\mathrm{B}}[k]||_2^2 && (20)\\ &= |x[k]|^2\big(\boldsymbol{q}^H[k]\boldsymbol{R}_{\mathrm{B}}[k]\boldsymbol{q}[k]-2\boldsymbol{q}^H[k]\boldsymbol{r}_{\mathrm{B}}[k]+\sigma_d^2[k]\big), && (21)\\ \mathcal{S}_{\mathrm{D}}(\boldsymbol{q}[k]) &= ||\mathbf{0}_{M_{\mathrm{D}}}-\boldsymbol{p}_{\mathrm{D}}[k]||_2^2 && (22)\\ &= |x[k]|^2\boldsymbol{q}^H[k]\boldsymbol{R}_{\mathrm{D}}[k]\boldsymbol{q}[k], && (23)\end{aligned}$$

where $||\cdot||_2$ is the ℓ_2-norm,

$$\begin{aligned}\boldsymbol{r}_{\mathrm{B}}[k] &= \boldsymbol{H}_{\mathrm{B}}^H[k]\boldsymbol{h}_z[k] \in \mathbb{C}^{L\times 1}\ , && (24)\\ \sigma_d^2[k] &= \boldsymbol{h}_z^H[k]\boldsymbol{h}_z[k]\ , && (25)\end{aligned}$$

where $\boldsymbol{r}_{\mathrm{B}}[k]$ is the spatial correlation vector. We refer to $\mathcal{S}_{\mathrm{B}}(\boldsymbol{q}[k])$ and $\mathcal{S}_{\mathrm{D}}(\boldsymbol{q}[k])$ as the signal distortion energy (SDE) in the bright zone and the residual energy (RE) in the dark zone, respectively.

Typically, minimizing SDE $\mathcal{S}_{\mathrm{B}}(\boldsymbol{q}[k])$ and/or RE $\mathcal{S}_{\mathrm{D}}(\boldsymbol{q}[k])$ is investigated in both the aspects of signal enhancement and sound zone control. This process can be implemented by posing a combined cost function of $\mathcal{S}_{\mathrm{B}}(\boldsymbol{q}[k])$ and $\mathcal{S}_{\mathrm{D}}(\boldsymbol{q}[k])$, i.e.,

$$\underset{\boldsymbol{q}[k]}{\arg\min}\ \Big(\mathcal{S}_{\mathrm{B}}(\boldsymbol{q}[k])+\mu\mathcal{S}_{\mathrm{D}}(\boldsymbol{q}[k])\Big)\ , \tag{26}$$

where μ is a user parameter that is used to weigh the importance between minimizing SDE and suppressing RE. It should be noted that the solution of (26) is identical to the solution obtained in [18] and in [32] if it is solved in a least squares manner in the frequency-domain and in the time-domain, respectively. Depending on the constraint that we consider, the physical meaning of μ changes, and the control filter gives a different performance. We return to this in Sec. IV.

D. Performance evaluation metrics

Inspired by the literature, e.g., in [11], [18], [49], we define the following metrics.

1) Prior acoustic contrast (iAC): We first measure the acoustic contrast before applying any filter, i.e.,

$$\gamma_{\mathrm{iAC}}[k] = \frac{\mathbf{1}_L^H\boldsymbol{R}_{\mathrm{B}}[k]\mathbf{1}_L}{\mathbf{1}_L^H\boldsymbol{R}_{\mathrm{D}}[k]\mathbf{1}_L}\ , \tag{27}$$

where $\mathbf{1}_L$ is an all-ones vector of length L. This is similar to the input SNR (iSNR) in signal enhancement.

2) Posterior acoustic contrast (oAC): The energy ratio between the bright and dark zones after being controlled can be defined as

$$\gamma_{\mathrm{oAC}}(\boldsymbol{q}[k]) = \frac{\boldsymbol{q}^H[k]\boldsymbol{R}_{\mathrm{B}}[k]\boldsymbol{q}[k]}{\boldsymbol{q}^H[k]\boldsymbol{R}_{\mathrm{D}}[k]\boldsymbol{q}[k]}\ , \tag{28}$$

which is typically called the acoustic contrast, as defined in [11]. Hence, the posterior acoustic contrast is denoted here simply as the acoustic contrast, unless otherwise specified. We can easily expect that oAC should be larger than iAC, i.e., $\gamma_{\mathrm{oAC}}(\boldsymbol{q}[k]) > \gamma_{\mathrm{iAC}}[k]$.

3) Normalized signal distortion energy in the bright zone (nSDE): We can measure the difference between the desired and reproduced sound fields using the signal distortion normalized by the energy of the desired sound field in the bright zone, i.e.,

$$\eta_{\mathrm{B}}(\boldsymbol{q}[k]) = \frac{\mathcal{S}_{\mathrm{B}}(\boldsymbol{q}[k])}{\boldsymbol{d}_{\mathrm{B}}^{H}[k]\boldsymbol{d}_{\mathrm{B}}[k]} = \frac{\sigma_d^2[k] + \boldsymbol{q}^H[k]\boldsymbol{R}_{\mathrm{B}}[k]\boldsymbol{q}[k] - 2\boldsymbol{q}^H[k]\boldsymbol{r}_{\mathrm{B}}[k]}{\sigma_d^2[k]} \ . \quad (29)$$

Note that if the reproduced sound field is close to the desired sound field, then this value becomes small, and ideally 0.

4) Normalized residual energy in the dark zone (nRE): This value is defined as a ratio between the energy in the dark zone before and after being controlled. We can write nRE as

$$\eta_{\mathrm{D}}(\boldsymbol{q}[k]) = \frac{\boldsymbol{q}^H[k]\boldsymbol{R}_{\mathrm{D}}[k]\boldsymbol{q}[k]}{\mathbf{1}_L^H\boldsymbol{R}_{\mathrm{D}}[k]\mathbf{1}_L} \ , \quad (30)$$

and apparently $\eta_{\mathrm{D}}(\boldsymbol{q}[k]) < 1$, which is similar to the noise reduction factor in signal enhancement.

5) A relationship between MSE criteria and the performance evaluation metrics: We can measure an MSE for the total field $\mathcal{S}_{\mathrm{F}}(\boldsymbol{q}[k])$, i.e.,

$$\mathcal{S}_{\mathrm{F}}(\boldsymbol{q}[k]) = ||\boldsymbol{d}_{\mathrm{F}}[k] - \boldsymbol{p}_{\mathrm{F}}[k]||_2^2 = \mathcal{S}_{\mathrm{B}}(\boldsymbol{q}[k]) + \mathcal{S}_{\mathrm{D}}(\boldsymbol{q}[k]) \ , \quad (31)$$

and if we plug (28) and (29) into (31), then we finally obtain

$$\mathcal{S}_{\mathrm{F}}(\boldsymbol{q}[k]) = |x[k]|^2\boldsymbol{q}^H[k]\boldsymbol{R}_{\mathrm{D}}[k]\boldsymbol{q}[k]\times \left(1 + \mathrm{oAC}(\boldsymbol{q}[k]) \times \eta_{\mathrm{B}}(\boldsymbol{q}[k]) \times \frac{\sigma_d^2[k]}{\boldsymbol{q}^H[k]\boldsymbol{R}_{\mathrm{B}}[k]\boldsymbol{q}[k]}\right) \ . \quad (32)$$

Now, we can see that the MSEs are highly related to oAC and nSDE. Note that the third term can be seen as the energy ratio between the desired and reproduced sound fields in the bright zone.

As alluded to earlier, we now introduce two subspace approaches in the DFT-domain to deal with (26).

III. VAST IN THE DFT-DOMAIN

In this section, subspace approaches for sound zone control are formulated in the DFT-domain.

A. Narrowband approach

It is well-known that we can jointly diagonalize two Hermitian matrices $\boldsymbol{R}_{\mathrm{B}}[k]$ and $\boldsymbol{R}_{\mathrm{D}}[k]$ such that

$$\boldsymbol{U}_L^H[k]\boldsymbol{R}_{\mathrm{B}}[k]\boldsymbol{U}_L[k] = \boldsymbol{\Lambda}_L[k] \ , \quad (33)$$

$$\boldsymbol{U}_L^H[k]\boldsymbol{R}_{\mathrm{D}}[k]\boldsymbol{U}_L[k] = \boldsymbol{I}_L \ , \quad (34)$$

where $\boldsymbol{U}_L[k]$ is an $L\times L$ nonsingular matrix, $\boldsymbol{\Lambda}_L[k]$ is an $L\times L$ diagonal matrix whose diagonal elements are nonnegative, and $\boldsymbol{I}_L$ is the $L \times L$ identity matrix [50]. Furthermore, $\boldsymbol{U}_L[k]$ and $\boldsymbol{\Lambda}_L[k]$ are the eigenvector and eigenvalue matrices of the generalized eigenvalue problem $\boldsymbol{R}_{\mathrm{D}}^{-1}[k]\boldsymbol{R}_{\mathrm{B}}[k]$ and more specifically

$$\boldsymbol{R}_{\mathrm{D}}^{-1}[k]\boldsymbol{R}_{\mathrm{B}}[k]\boldsymbol{U}_L[k] = \boldsymbol{U}_L[k]\boldsymbol{\Lambda}_L[k] \ , \quad (35)$$

where

$$\boldsymbol{\Lambda}_L[k] = \mathrm{diag}(\lambda_1[k], \cdots, \lambda_L[k]) \in \mathbb{R}^{L\times L} \ , \quad (36)$$

$$\boldsymbol{U}_L[k] = \left[\boldsymbol{u}_1[k], \ \cdots, \ \boldsymbol{u}_L[k]\right] \in \mathbb{C}^{L\times L} \ , \quad (37)$$

where $\lambda_v[k]$ and $\boldsymbol{u}_v[k]$ are the vth eigenvalue and eigenvector, respectively, and $1 \le v \le L$. Typically, the eigenvalues are sorted in descending order, which is $\lambda_1[k] \ge \cdots \ge \lambda_L[k] \ge 0$, and the corresponding eigenvectors follow the same order. Note that the rank of $\boldsymbol{R}_{\mathrm{D}}[k]$ is $\mathrm{rank}(\boldsymbol{R}_{\mathrm{D}}[k]) \le \min(M_{\mathrm{D}}, L)$. If the number of the control points is larger than that of the loudspeakers, i.e., $M_{\mathrm{D}} \ge L$, then $\mathrm{rank}(\boldsymbol{R}_{\mathrm{D}}[k]) \le L$. The spatial correlation matrix $\boldsymbol{R}_{\mathrm{D}}[k]$ is a full rank matrix if $\mathrm{rank}(\boldsymbol{H}_{\mathrm{D}}[k]) = L$. Alternatively, depending on the target frequency and/or the system geometry, a regularization such as the Tikhonov regularization can be included to guarantee that $\boldsymbol{R}_{\mathrm{D}}[k]$ has full rank so that its inverse can be computed [25], [51]. In contrast, $\boldsymbol{R}_{\mathrm{B}}[k]$ does not have to be a positive definite matrix. Rather, $\boldsymbol{R}_{\mathrm{B}}[k]$ can have any rank from 1 to L, which is less strict than $\boldsymbol{R}_{\mathrm{D}}[k]$. If $\mathrm{rank}(\boldsymbol{R}_{\mathrm{D}}[k]) = Q$, then the last $L - Q$ eigenvalues in $\boldsymbol{\Lambda}_L[k]$ are equal to 0.

Because any vector can be represented as a linear combination of the columns of a nonsingular matrix, we can express the control filter $\boldsymbol{q}[k]$ using $\boldsymbol{U}_L[k]$ as

$$\boldsymbol{q}[k] = \boldsymbol{U}_L[k]\boldsymbol{a}_L[k] \ , \quad (38)$$

where $\boldsymbol{a}_L[k]$ is a coefficient vector of length L. If we insert (38) into (21) and (23), then we obtain

$$\mathcal{S}_{\mathrm{B}}(\boldsymbol{U}_L[k]\boldsymbol{a}_L[k]) = |x[k]|^2\big(\boldsymbol{a}_L^H[k]\boldsymbol{\Lambda}_L[k]\boldsymbol{a}_L[k] - 2\boldsymbol{a}_L^H[k]\boldsymbol{U}_L^H[k]\boldsymbol{r}_{\mathrm{B}}[k] + \sigma_d^2[k]\big), \quad (39)$$

$$\mathcal{S}_{\mathrm{D}}(\boldsymbol{U}_L[k]\boldsymbol{a}_L[k]) = |x[k]|^2|\boldsymbol{a}_L[k]|^2 \ . \quad (40)$$

Interestingly, RE $\mathcal{S}_{\mathrm{D}}(\boldsymbol{U}_L[k]\boldsymbol{a}_L[k])$ depends only on the input signal $x[k]$ and on the coefficient vector $\boldsymbol{a}_L[k]$. Making the last $L - V$ coefficients in $\boldsymbol{a}_L[k]$ be 0 reduces the RE directly. We can obtain such a benefit using the so-called subspace-based approach. In other words, we can approximate the control filter $\boldsymbol{q}[k]$ using a V-rank approximation of $\boldsymbol{q}[k]$, i.e.,

$$\boldsymbol{q}[k] \approx \boldsymbol{U}_V[k]\boldsymbol{a}_V[k] \ , \quad (41)$$

where $1 \le V \le L$. If we re-write (39) and (40) using (41), then we obtain

$$\mathcal{S}_{\mathrm{B}}(\boldsymbol{U}_V[k]\boldsymbol{a}_V[k]) = |x[k]|^2\big(\boldsymbol{a}_V^H[k]\boldsymbol{\Lambda}_V[k]\boldsymbol{a}_V[k] - 2\boldsymbol{a}_V^H[k]\boldsymbol{U}_V^H[k]\boldsymbol{r}_{\mathrm{B}}[k] + \sigma_d^2[k]\big), \quad (42)$$

$$\mathcal{S}_{\mathrm{D}}(\boldsymbol{U}_V[k]\boldsymbol{a}_V[k]) = |x[k]|^2\boldsymbol{a}_V^H[k]\boldsymbol{a}_V[k] \ . \quad (43)$$

Here, we design the control filter $\boldsymbol{q}[k]$ for frequency bin k, which is a traditional approach in the frequency-domain, e.g., [11], [13], [18]. If we directly pose the optimization problem that seeks the control filters to minimize $\mathcal{S}_{\mathrm{F}}(\boldsymbol{q}[k])$, which is the case of $\mu = 1$ and $V = L$ in (26), then this approach yields the PM solution [13].

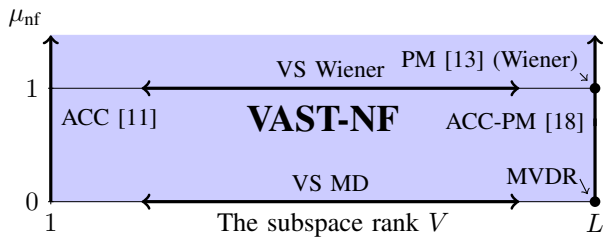


Fig. 2. A (V, μ_{nf}) plane that illustrates the relationship between VAST-NF and existing methods, i.e., ACC [11], PM [13], and ACC-PM [18].

If we plug (41), (42), and (43) into (26), then we obtain the following cost function $\mathcal{L}_{\text{nf}}(\boldsymbol{U}_V[k]\boldsymbol{a}_V[k])$ such that

$$\begin{aligned}\mathcal{L}_{\text{nf}}(\boldsymbol{U}_V[k]\boldsymbol{a}_V[k]) &= \mathcal{S}_{\text{B}}(\boldsymbol{U}_V[k]\boldsymbol{a}_V[k]) + \mu_{\text{nf}}\mathcal{S}_{\text{D}}(\boldsymbol{U}_V[k]\boldsymbol{a}_V[k]) \\ &= |x[k]|^2\Big(\boldsymbol{a}_V^H[k]\boldsymbol{\Lambda}_V[k]\boldsymbol{a}_V[k] \\ &\quad - 2\boldsymbol{a}_V^H[k]\boldsymbol{U}_V^H[k]\boldsymbol{r}_{\text{B}}[k] + \sigma_d^2[k]\Big) \\ &\quad + \mu_{\text{nf}}[k]|x[k]|^2\boldsymbol{a}_V^H[k]\boldsymbol{a}_V[k] \ , \end{aligned} \tag{44}$$

where $\mu_{\text{nf}}[k]$ is the frequency-dependent Lagrange multiplier for the narrowband approach. Setting to 0 the derivative of (44) with respect to $\boldsymbol{a}_V[k]$ allows us to obtain the optimal coefficient vector, which is given by

$$\boldsymbol{a}_V[k] = \left(\boldsymbol{\Lambda}_V[k] + \mu_{\text{nf}}[k]\boldsymbol{I}_V\right)^{-1}\boldsymbol{U}_V^H[k]\boldsymbol{r}_{\text{B}}[k] \ , \tag{45}$$

then the control filter, referred to as the narrowband variable span trade-off (VAST-NF) filter, yields

$$\begin{aligned}\boldsymbol{q}_{\text{VAST-NF}}[k] &= \boldsymbol{U}_V[k]\boldsymbol{a}_V[k] \\ &= \boldsymbol{U}_V[k]\left(\boldsymbol{\Lambda}_V[k] + \mu_{\text{nf}}[k]\boldsymbol{I}_V\right)^{-1}\boldsymbol{U}_V^H[k]\boldsymbol{r}_{\text{B}}[k] \\ &= \sum_{v=1}^{V}\frac{\boldsymbol{u}_v^H[k]\boldsymbol{r}_{\text{B}}[k]}{\mu_{\text{nf}}[k] + \lambda_v[k]}\boldsymbol{u}_v[k] \ , \end{aligned} \tag{46}$$

for $1 \leq V \leq L$ and $\mu_{\text{nf}}[k] \geq 0$.

Interestingly, VAST-NF reduces to existing sound zone control methods in special cases by varying the user parameters. For $V = 1$ and $V = L$, VAST-NF reduces to ACC in [11] and ACC-PM in [18], respectively; furthermore, it reduces to the minimum variance distortionless response (MVDR) filter for $V = L$ and $\mu = 0$ [39]. For $1 < V < L$ with $\mu = 0$ and $\mu = 1$, it reduces to the variable span (VS) Wiener and VS minimum distortion (MD), respectively[3]. The details are summarized in Fig. 2 and analyzed in Sec. IV.

B. Broadband approach

For K frequency bins, one typical way to extend a narrowband approach to the case of broadband is to solve each frequency bin separately. However, we tackle one general problem that considers K frequency bins at the same time, which we refer to as the broadband approach, instead of K independent problems.

[3]The terminologies – VS MD, VS Wiener, MVDR – are inspired by the usage in [37], [49] and speech enhancement in general.

As a broadband sound zone control problem, we model the reproduced and desired sound fields using (9) and (15) as

$$\underline{\boldsymbol{p}}_{\text{C}} = \left[\boldsymbol{p}_{\text{C}}^T[0], \quad \cdots, \quad \boldsymbol{p}_{\text{C}}^T[K-1]\right]^T \in \mathbb{C}^{M_{\text{C}}K\times 1} \ , \tag{47}$$

$$= \underline{\boldsymbol{H}}_{\text{C}}\underline{\boldsymbol{q}} \quad \text{for C} \in \{\text{B, D}\} \ , \tag{48}$$

$$\underline{\boldsymbol{d}}_{\text{B}} = \left[\boldsymbol{d}_{\text{B}}^T[0], \quad \cdots, \quad \boldsymbol{d}_{\text{B}}^T[K-1]\right]^T \in \mathbb{C}^{M_{\text{B}}K\times 1} \ , \tag{49}$$

where

$$\underline{\boldsymbol{H}}_{\text{C}} = \text{blkdiag}(x[0]\boldsymbol{H}_{\text{C}}[0], \cdots, x[K-1]\boldsymbol{H}_{\text{C}}[K-1]) \ , \tag{50}$$

$$\underline{\boldsymbol{q}} = \left[\boldsymbol{q}^T[0], \quad \cdots, \quad \boldsymbol{q}^T[K-1]\right]^T \in \mathbb{C}^{LK\times 1} \ . \tag{51}$$

Because all K frequency bins are taken into account in $\underline{\boldsymbol{p}}_{\text{C}}$ and $\underline{\boldsymbol{d}}_{\text{B}}$, the frequency index k is no longer used. Note that the number of frequency bins K is $K = J/2+1$ if J is even, otherwise $K = (J+1)/2$. Now, we can consider the two MSEs for the broadband scenario as

$$\mathcal{S}_{\text{B}}(\underline{\boldsymbol{q}}) = ||\underline{\boldsymbol{d}}_{\text{B}} - \underline{\boldsymbol{p}}_{\text{B}}||_2^2 \tag{52}$$

$$= \hat{\sigma}_d^2 - 2\underline{\boldsymbol{q}}^H\underline{\boldsymbol{r}}_{\text{B}} + \underline{\boldsymbol{q}}^H\underline{\boldsymbol{R}}_{\text{B}}\underline{\boldsymbol{q}} \tag{53}$$

$$= \hat{\sigma}_d^2 - 2\underline{\boldsymbol{q}}^H\underline{\boldsymbol{X}}\underline{\boldsymbol{r}}'_{\text{B}} + \underline{\boldsymbol{q}}^H\underline{\boldsymbol{X}}\underline{\boldsymbol{R}}'_{\text{B}}\underline{\boldsymbol{q}} \ , \tag{54}$$

$$\mathcal{S}_{\text{D}}(\underline{\boldsymbol{q}}) = ||\boldsymbol{0}_{M_{\text{D}}K} - \underline{\boldsymbol{p}}_{\text{D}}||_2^2 \tag{55}$$

$$= \underline{\boldsymbol{q}}^H\underline{\boldsymbol{R}}_{\text{D}}\underline{\boldsymbol{q}} \tag{56}$$

$$= \underline{\boldsymbol{q}}^H\underline{\boldsymbol{X}}\underline{\boldsymbol{R}}'_{\text{D}}\underline{\boldsymbol{q}} \ , \tag{57}$$

where

$$\underline{\boldsymbol{R}}_{\text{C}} = \underline{\boldsymbol{X}}\underline{\boldsymbol{R}}'_{\text{C}} \quad \text{for C} \in \{\text{B, D}\} \ , \tag{58}$$

$$\underline{\boldsymbol{R}}'_{\text{C}} = \text{blkdiag}\left(\boldsymbol{R}_{\text{C}}[0], \cdots, \boldsymbol{R}_{\text{C}}[K-1]\right) \ , \tag{59}$$

$$\boldsymbol{R}_{\text{C}}[k] = \boldsymbol{H}_{\text{C}}^H[k]\boldsymbol{H}_{\text{C}}[k] \ , \tag{60}$$

$$\underline{\boldsymbol{X}} = \boldsymbol{X} \otimes \boldsymbol{I}_L \ , \tag{61}$$

$$\boldsymbol{X} = \text{diag}\left(|x[0]|^2, \cdots, |x[K-1]|^2\right) \ , \tag{62}$$

$$\underline{\boldsymbol{r}}_{\text{B}} = \underline{\boldsymbol{X}}\underline{\boldsymbol{r}}'_{\text{B}} \ , \tag{63}$$

$$\underline{\boldsymbol{r}}'_{\text{B}} = \left[\boldsymbol{r}_{\text{B}}^T[0], \quad \cdots, \quad \boldsymbol{r}_{\text{B}}^T[K-1]\right]^T \ , \tag{64}$$

$$\hat{\sigma}^2 = \underline{\boldsymbol{d}}_{\text{B}}^H\underline{\boldsymbol{d}}_{\text{B}} \ , \tag{65}$$

where $\underline{\boldsymbol{r}}_{\text{B}}$ is the cross-correlation vector in the broadband approach.

Here, we consider the joint diagonalization of two $LK\times LK$ Hermitian matrices $\underline{\boldsymbol{R}}_{\text{B}}$ and $\underline{\boldsymbol{R}}_{\text{D}}$, i.e.,

$$\underline{\boldsymbol{U}}_{LK}^H\underline{\boldsymbol{R}}_{\text{B}}\underline{\boldsymbol{U}}_{LK} = \underline{\boldsymbol{\Lambda}}_{LK} \ , \quad \underline{\boldsymbol{U}}_{LK}^H\underline{\boldsymbol{R}}_{\text{D}}\underline{\boldsymbol{U}}_{LK} = \boldsymbol{I}_{LK} \ , \tag{66}$$

where $\underline{\boldsymbol{\Lambda}}_{LK}$ and $\underline{\boldsymbol{U}}_{LK}$ are the eigenvalues and eigenvectors of the generalized eigenvalue problem of

$$\underline{\boldsymbol{R}}_{\text{D}}^{-1}\underline{\boldsymbol{R}}_{\text{B}}\underline{\boldsymbol{U}}_{LK} = \underline{\boldsymbol{U}}_{LK}\underline{\boldsymbol{\Lambda}}_{LK} \ , \tag{67}$$

and $\underline{\boldsymbol{\Lambda}}_{LK}$ can be sorted in descending order globally, and $\underline{\boldsymbol{U}}_{LK}$ is sorted in the same order.

Next, we assume that $\underline{\boldsymbol{\Lambda}}_{LK}$ and $\underline{\boldsymbol{U}}_{LK}$ are globally sorted in descending order. Interestingly, each eigenvector only contains one frequency bin information in this approach as $\underline{\boldsymbol{R}}_{\text{D}}^{-1}\underline{\boldsymbol{R}}_{\text{B}}$ is also block diagonal and can be computed block-wise, which means that each eigenvector includes at least $(L-1)K$ zeros. This is due to the fact that the eigenvectors of $\underline{\boldsymbol{R}}_{\text{D}}^{-1}\underline{\boldsymbol{R}}_{\text{B}}$ can be

obtained by padding zeros to the eigenvectors obtained from each block matrix in it.

Here, we apply the V-rank approximation to the control filter $\underline{\boldsymbol{q}}$, as in the previous section, i.e.,

$$\underline{\boldsymbol{q}} \approx \underline{\boldsymbol{U}}_V \underline{\boldsymbol{a}}_V \ , \tag{68}$$

where $1 \leq V \leq LK$. Then, we can obtain

$$\mathcal{S}_{\mathrm{B}}(\underline{\boldsymbol{U}}_V \underline{\boldsymbol{a}}_V) = \hat{\sigma}_d^2 - 2\underline{\boldsymbol{a}}_V^H \underline{\boldsymbol{U}}_V^H \underline{\boldsymbol{r}}_{\mathrm{B}} + \underline{\boldsymbol{a}}_V^H \underline{\boldsymbol{\Lambda}}_V \underline{\boldsymbol{a}}_V \ , \tag{69}$$

$$\mathcal{S}_{\mathrm{D}}(\underline{\boldsymbol{U}}_V \underline{\boldsymbol{a}}_V) = \underline{\boldsymbol{a}}_V^H \underline{\boldsymbol{a}}_V \ . \tag{70}$$

If we pose an optimization as in (26), then the broadband VAST (VAST-BF) control filter $\underline{\boldsymbol{q}}_{\text{VAST-BF}}$ is given by

$$\begin{aligned} \underline{\boldsymbol{q}}_{\text{VAST-BF}} &= \arg\min_{\underline{\boldsymbol{a}}_V} \left(\mathcal{S}_{\mathrm{B}}(\underline{\boldsymbol{U}}_V \underline{\boldsymbol{a}}_V) + \mu_{\text{bf}} \mathcal{S}_{\mathrm{D}}(\underline{\boldsymbol{U}}_V \underline{\boldsymbol{a}}_V) \right) \\ &= \sum_{v=1}^{V} \frac{\underline{\boldsymbol{u}}_v^H \underline{\boldsymbol{r}}_{\mathrm{B}}}{\mu_{\text{bf}} + \lambda_v} \underline{\boldsymbol{u}}_v \ , \end{aligned} \tag{71}$$

where $1 \leq V \leq LK$, μ_{bf} is the Lagrange multiplier for the broadband approach, and λ_v and $\underline{\boldsymbol{u}}_v$ are the vth generalized eigenvalue and eigenvector, respectively.

It should also be noted that we can obtain the time-domain VAST control filter (VAST-T) if we follow the same procedure as that explained in this section. To this end, the control filter $\hat{\underline{\boldsymbol{q}}}_{\text{VAST-T}}$ is given by

$$\hat{\underline{\boldsymbol{q}}}_{\text{VAST-T}} = \sum_{v=1}^{V} \frac{\hat{\underline{\boldsymbol{u}}}_v^T \hat{\underline{\boldsymbol{r}}}_{\mathrm{B}}}{\hat{\mu}_{\mathrm{t}} + \hat{\lambda}_v} \hat{\underline{\boldsymbol{u}}}_v \ , \tag{72}$$

where $1 \leq V \leq LJ$, J is the length of the control filter, $\hat{\mu}_{\mathrm{t}}$ is the Lagrange multiplier for the time-domain approach, $\hat{\underline{\boldsymbol{u}}}_v$ and $\hat{\lambda}_v$ are the vth eigenvector and eigenvalue, respectively, and $\hat{\underline{\boldsymbol{r}}}_{\mathrm{B}}$ is the spatial cross-correlation vector in the time-domain. See [39] for more detail on the derivations.

It is worth noting the following: 1) in the broadband approaches, i.e., VAST-BF and VAST-T, we have a single constraint, and not K independent constraints as in the narrowband approach, VAST-NF. 2) The VAST framework gives a high degree of flexibility in the control filter design. In other words, we can obtain VAST-NF as described in Sec. III-A, VAST-BF as described in Sec. III-B, and VAST-T as explained in [39], respectively, depending on the domain in which the reproduced and desired sound fields are formulated. 3) The performance evaluation metrics of the narrowband approach, which is described in Sec. II-D can also be used for VAST-BF and VAST-T without needing to significantly modify these methods.

IV. Analysis and Discussion

In the previous section, two methods in the DFT-domain for sound zone control were proposed. In this section, we analyze the proposed methods. First, we investigate the lower- and upper-bounds of the MSEs as well as the performance metrics that were previously proposed in the time-domain [42], [45]. This approach allows us to confirm the core property of the proposed methods, i.e., the trade-off between oAC and SDE. Second, we examine different types of constraints to find the optimal μ for the proposed methods. Third, we consider the computational complexity of the time-domain approach and the frequency-domain approaches including the proposed methods. Finally, we discuss the similarity and difference between the two proposed methods. To this end, we can fully understand the proposed methods. It should be noted that this section is based on the narrowband approach, but this method can also be applied to not only the broadband approach but also the time-domain approach without the loss of generality.

A. Performance boundaries for the MSE criteria

In this subsection, we consider the bounds for the MSEs. These bounds are determined by the two user parameters, i.e., V and μ_{nf}. We consider the bounds as a function of V first. Plugging (45) into (42), (43), and (44), respectively, leads us to

$$\begin{aligned} &\mathcal{S}_{\mathrm{B}}(\boldsymbol{U}_V[k]\boldsymbol{a}_V[k]) \\ &= |x[k]|^2 \left(\sigma_d^2[k] - \sum_{v=1}^{V} \frac{2\mu_{\text{nf}}[k] + \lambda_v[k]}{(\mu_{\text{nf}}[k] + \lambda_v[k])^2} |\boldsymbol{u}_v^H[k]\boldsymbol{r}_{\mathrm{B}}[k]|^2 \right) , \end{aligned} \tag{73}$$

$$\mathcal{S}_{\mathrm{D}}(\boldsymbol{U}_V[k]\boldsymbol{a}_V[k]) = |x[k]|^2 \sum_{v=1}^{V} \frac{|\boldsymbol{u}_v^H[k]\boldsymbol{r}_{\mathrm{B}}[k]|^2}{(\mu_{\text{nf}}[k] + \lambda_v[k])^2} \ , \tag{74}$$

and

$$\begin{aligned} &\mathcal{L}_{\text{nf}}(\boldsymbol{U}_V[k]\boldsymbol{a}_V[k]) \\ &\quad = \mathcal{S}_{\mathrm{B}}(\boldsymbol{U}_V[k]\boldsymbol{a}_V[k]) + \mu_{\text{nf}}[k] \mathcal{S}_{\mathrm{D}}(\boldsymbol{U}_V[k]\boldsymbol{a}_V[k]) \\ &\quad = |x[k]|^2 \left(\sigma_d^2[k] - \sum_{v=1}^{V} \frac{|\boldsymbol{u}_v^H[k]\boldsymbol{r}_{\mathrm{B}}[k]|^2}{\mu_{\text{nf}}[k] + \lambda_v[k]} \right) . \end{aligned} \tag{75}$$

As V increases with a fixed μ_{nf}, SDE monotonically decreases, whereas RE monotonically increases. However, fortunately, the cost function $\mathcal{L}_{\text{nf}}$ finally decreases.

Now, we consider the bounds of SDE, RE, and the cost function $\mathcal{L}_{\text{nf}}$ as a function of μ_{nf}. Specifically, we determine the upper and lower bounds for the two extreme cases, $\mu_{\text{nf}} = 0$ and $\mu_{\text{nf}} \to \infty$, respectively, i.e.,

$$\begin{aligned} &\mathcal{S}_{\mathrm{B}}(\boldsymbol{U}_V[k]\boldsymbol{a}_V[k]) \\ &\quad = \begin{cases} |x[k]|^2 \left(\sigma_d^2[k] - \sum_{v=1}^{V} \xi_v[k] \right) , & \text{for } \mu_{\text{nf}}[k] = 0 \\ |x[k]|^2 \sigma_d^2[k], & \text{for } \mu_{\text{nf}}[k] \to \infty \ , \end{cases} \end{aligned} \tag{76}$$

$$\begin{aligned} &\mathcal{S}_{\mathrm{D}}(\boldsymbol{U}_V[k]\boldsymbol{a}_V[k]) \\ &\quad = \begin{cases} |x[k]|^2 \sum_{v=1}^{V} \lambda_v^{-1}[k] \xi_v[k], & \text{for } \mu_{\text{nf}}[k] = 0 \\ 0, & \text{for } \mu_{\text{nf}}[k] \to \infty \ , \end{cases} \end{aligned} \tag{77}$$

where $\xi_v[k] = \lambda_v^{-1}[k] |\boldsymbol{u}_v^H[k]\boldsymbol{r}_{\mathrm{B}}[k]|^2$, and the cost function $\mathcal{L}_{\text{nf}}$ has the same bounds as SDE. Note that the largest SDE is made when $\mu_{\text{nf}} \to \infty$, whereas the largest RE is obtained when $V = L$ and $\mu_{\text{nf}} = 0$, and vice versa. Therefore, we state the following conclusions. SDE and the cost function $\mathcal{L}_{\text{nf}}$ decrease with increasing V and decreasing μ_{nf}, whereas RE behaves in the opposite manner.

B. Performance boundaries for the performance metrics

Now, we analyze the bounds for the performance metrics with respect to the two user parameters. Using (46), oAC can be re-written as

$$\gamma_{\text{oAC}}(\boldsymbol{U}_V[k]\boldsymbol{a}_V[k]) = \frac{\boldsymbol{a}_V^H[k]\boldsymbol{\Lambda}_V[k]\boldsymbol{a}_V[k]}{\boldsymbol{a}_V^H[k]\boldsymbol{a}_V[k]} \tag{78}$$

$$= \frac{\sum_{v=1}^{V} \frac{\lambda_v[k]|\boldsymbol{u}_v^H[k]\boldsymbol{r}_{\text{B}}[k]|^2}{(\mu_{\text{nf}}[k]+\lambda_v[k])^2}}{\sum_{v=1}^{V} \frac{|\boldsymbol{u}_v^H[k]\boldsymbol{r}_{\text{B}}[k]|^2}{(\mu_{\text{nf}}[k]+\lambda_v[k])^2}} , \tag{79}$$

and oAC decreases with increasing V, as shown in [39], [41], and oAC is always maximized when $V = 1$ regardless of μ_{nf}, i.e., $\gamma_{\text{oAC}}(\boldsymbol{U}_1[k]\boldsymbol{a}_1[k]) = \lambda_1[k]$. In contrast, the bounds of oAC with respect to μ_{nf} are expressed as

$$\gamma_{\text{oAC}}(\boldsymbol{U}_V[k]\boldsymbol{a}_V[k]) = \begin{cases} \frac{\sum_{v=1}^{V}\xi_v[k]}{\sum_{v=1}^{V}\lambda_v^{-1}[k]\xi_v[k]}, & \text{for } \mu_{\text{nf}}[k] = 0 \\ \frac{\sum_{v=1}^{V}\lambda_v^{2}[k]\xi_v[k]}{\sum_{v=1}^{V}\lambda_v[k]\xi_v[k]}, & \text{for } \mu_{\text{nf}}[k] \to \infty \end{cases} , \tag{80}$$

and oAC decreases with decreasing μ_{nf}. Because both oAC and SDE increase with decreasing V and/or increasing μ_{nf}, we obtain a high oAC at the cost of a large SDE; therefore, we can observe a trade-off between oAC and SDE.

In the same manner, the bounds of nSDE can be established as follows. These are easily obtained from (73) with respect to V and (76) with respect to μ_{nf}, respectively, i.e.,

$$\eta_{\text{B}}(\boldsymbol{U}_V[k]\boldsymbol{a}_V[k]) = \frac{\mathcal{S}_{\text{B}}(\boldsymbol{U}_V[k]\boldsymbol{a}_V[k])}{|\boldsymbol{d}_{\text{B}}[k]|^2} \tag{81}$$

$$= \begin{cases} 1 - \sigma_d^{-2}[k]\sum_{v=1}^{V}\xi_v[k], & \text{for } \mu_{\text{nf}}[k] = 0 \\ 1, & \text{for } \mu_{\text{nf}}[k] \to \infty \end{cases} . \tag{82}$$

Because nSDE is a signal-independent performance metric that is normalized by the energy of the desired sound field in the bright zone $|\boldsymbol{d}_{\text{B}}[k]|^2$, regardless of V and μ_{nf}, nSDE is bounded by $0 \leq \eta_{\text{B}}(\boldsymbol{U}_V[k]\boldsymbol{a}_V[k]) \leq 1$.

Finally, the bounds of nRE can be computed by plugging (46) into (30) such that

$$\eta_{\text{D}}(\boldsymbol{q}[k]) = \frac{\sum_{v=1}^{V} \frac{|\boldsymbol{u}_v^H[k]\boldsymbol{r}_{\text{B}}[k]|^2}{(\mu_{\text{nf}}[k]+\lambda_v[k])^2}}{\mathbf{1}_L^H\boldsymbol{R}_{\text{D}}[k]\mathbf{1}_L} \tag{83}$$

$$= \begin{cases} \frac{\sum_{v=1}^{V}\lambda_v^{-1}[k]\xi_v[k]}{\mathbf{1}_L^H\boldsymbol{R}_{\text{D}}[k]\mathbf{1}_L}, & \text{for } \mu_{\text{nf}}[k] = 0 \\ 0, & \text{for } \mu_{\text{nf}}[k] \to \infty \end{cases} . \tag{84}$$

For the case of $\mu_{\text{nf}} \to \infty$, nRE converges to 0, which means that no energy remains in the dark zone. However, this is only valid when the magnitude of the control filter $|\boldsymbol{q}[k]|$ is almost equal to 0, which might only be meaningful in a heuristic purpose.

C. Optimization criteria

In the previous subsection, we investigated the performance boundaries of the proposed methods. Any value for the two user parameters $1 \leq V \leq L$ and $0 \leq \mu_{\text{nf}}$ can be used to obtain a certain performance from the proposed methods by manual tuning as in [39]–[41]; however, it is difficult to find a proper value of the desired RE on the constraint. Recently, various strategies for finding an optimal value using physically meaningful constraints in the context of sound zones were proposed in [45]. These strategies can also be applied to the methods proposed here.

In this subsection, we review the typical convex optimization problem in the context of sound zones, and then we investigate two types of constraints. Note that the convex optimization form in (26) does not change, while the solution space for μ_{nf} varies depending on which constraint is considered.

1) Constraint on RE: A constraint on RE is typical in the context of sound zones, e.g., [11], [18], [32], [39]; therefore, we can pose the optimization problem as follows:

$$\underset{\boldsymbol{q}[k]}{\text{minimize}}\ \mathcal{S}_{\text{B}}(\boldsymbol{q}[k]) \text{ subject to } \mathcal{S}_{\text{D}}(\boldsymbol{q}[k]) \leq \epsilon[k] , \tag{85}$$

where $\epsilon[k]$ is the desired amount of energy in the dark zone. Note that (44) is the Lagrangian associated with (85), and the optimal solution to (85) is the VAST-NF control filter in (46). The tuning parameter μ_{nf} in (46) is directly related to $\epsilon[k]$, because finding the optimal $\mu_{\text{nf}}^{\text{opt}}$ is required to fulfil the constraint. This optimal $\mu_{\text{nf}}^{\text{opt}}$ can be found by using one of the root finding methods, e.g., Newton's method, as in [16], [42]. However, as alluded to earlier in this subsection, it is often difficult to assign a proper value of $\epsilon[k]$ because this parameter varies depending on several factors, such as frequency and the geometry of the loudspeaker array.

2) Constraint on nRE: The challenge of selecting a proper $\epsilon[k]$ can be rather easily and intuitively resolved by posing the optimization problem with a constraint on nRE, i.e.,

$$\underset{\boldsymbol{q}[k]}{\text{minimize}}\ \mathcal{S}_{\text{B}}(\boldsymbol{q}[k]) \text{ subject to } \eta_{\text{D}}(\boldsymbol{q}[k]) \leq \epsilon[k] . \tag{86}$$

For example, $\epsilon[k] = -10$ dB is not a reasonable choice if RE is already $\mathcal{S}_{\text{D}}(\boldsymbol{q}[k]) < -10$ dB, whereas $\eta_{\text{D}}(\boldsymbol{q}[k]) < -10$ dB can be possible in most cases because the desired residual energy is 10 dB less than the initial energy in the dark zone, i.e., $\mathbf{1}_L^H\boldsymbol{R}_{\text{D}}[k]\mathbf{1}_L$.

From (86), a cost function similar to (44) can be obtained. Following the same procedure described in Sec. III-A, we can obtain the same control filter as (46), except for μ_{nf}, which yields $\mu_{\text{nf}}/\mathbf{1}_L^H\boldsymbol{R}_{\text{D}}[k]\mathbf{1}_L$. However, $\mu_{\text{nf}}^{\text{opt}}$ can be optimally computed according to the constraint by using any root finding method, as mentioned in the subsection above. The derivations are omitted due to their similarity.

3) Constraint on nSDE: Controlling nRE with an optimal value that corresponds to the constraint might be insufficient to guarantee that the reproduced sound field in the bright zone is as close to the desired sound field as possible. Hence, to tackle this issue, nSDE might be one of the constraints that we can consider. For the same reason as nRE, we use nSDE over SDE.

Next, we minimize RE with a constraint on nSDE; therefore, the optimization problem yields

$$\underset{\boldsymbol{q}[k]}{\text{minimize}}\ \mathcal{S}_{\text{D}}(\boldsymbol{q}[k]) \text{ subject to } \eta_{\text{B}}(\boldsymbol{q}[k]) \leq \epsilon[k] . \tag{87}$$

Again, a cost function similar to (44) from (87) can be established. This optimization problem can be solved by following

TABLE I
THE COMPUTATIONAL COMPLEXITY FOR THE CALCULATION OF SPATIAL STATISTICS AND CONTROL FILTERS

Method	Domain	Spatial statistics	Control filter
BACC-PM [32] VAST-T [39]	Time	$\mathcal{O}(MNL^2J^2)$	$\mathcal{O}(L^3J^3)$
ACC [11] PM [13] VAST-NF VAST-BF	Frequency	$\mathcal{O}(L^2K)$	$\mathcal{O}(L^3K)$

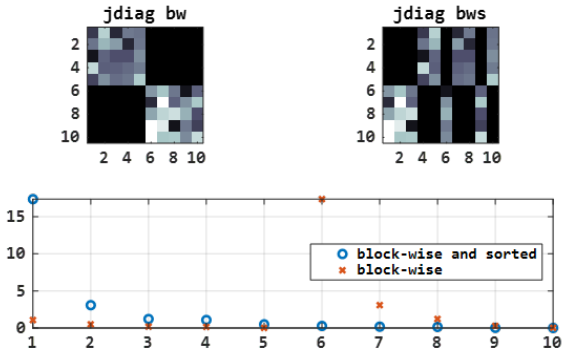


Fig. 3. An illustration of block-wise joint diagonalization for $\underline{\boldsymbol{U}}_{LK}$ with $L = 5$ and $K = 2$. The upper row shows the eigenvectors after the joint diagonalization is computed block-wise (in the upper left corner denoted as `jdiag bw`), and the upper right corner panel (in the upper right corner denoted as `jdiag bws`) shows the eigenvectors after joint diagonalization and global sorting according to the eigenvalues in the lower panel. Note that the larger and smaller values are white and black, respectively, while the values between are appropriately shaded. The panel in the lower row shows the sorted eigenvalues for the upper left panel (red cross) and for the upper right panel (blue circle), respectively.

Sec. III-A. Finally, $\mu_{\text{nf}}^{\text{opt}}$ can also be calculated by using any root finding method.

D. Computational complexity

In this section, we investigate the computational complexity of the proposed methods and several existing methods: ACC in [11], PM in [13], BACC-PM in [32], and VAST-T in [39]. As summarized in Table I, VAST-T and BACC-PM are the most computationally complex because they are time-domain approaches. Here, we can clearly observe that the time-domain approaches have a high degree of computational complexity to compute both the spatial statistics and the control filters compared to that of the frequency-domain approaches. As the length of the control filter J becomes longer, this difference becomes bigger. In practice, the number of frequency bins K is smaller than J, as alluded to earlier, the frequency-domain approaches demand even less computational complexity than that of the time-domain approaches. Note that VAST-BF can compute the joint digonalization block-wise, it has the same computational complexity as VAST-NF.

E. Discussion on VAST-NF and VAST-BF

In the previous subsections, the performance boundaries for the MSEs and the performance metrics were investigated. It is worth noting that the MSEs and the cost function are signal-dependent. However, each of the performance metrics is signal-independent. This result is due to the fact that the metrics are defined as a ratio between two quantities, and both quantities contain the contribution of the input signal, which cancels out. In (46), we can see that $\lambda_v[k]$, $\boldsymbol{u}_v[k]$, and $\boldsymbol{r}_{\text{B}}[k]$ are signal-independent. Therefore, the narrowband approach VAST-NF can be considered as signal-independent if V and μ_{nf} are tuned manually.

Now, if we consider the broadband approach, we can write the generalized eigenvalue problem of the kth block in $\underline{\boldsymbol{R}}_{\text{C}}$ in (58) as $\underline{\boldsymbol{R}}_{\text{C}}$ is block diagonal, i.e.,

$$|x[k]|^2\boldsymbol{R}_{\text{B}}[k]\boldsymbol{U}_L[k] = |x[k]|^2\boldsymbol{R}_{\text{D}}[k]\boldsymbol{U}_L[k]\boldsymbol{\Lambda}_L[k] \ . \tag{88}$$

If we compute the joint diagonalization from (88), then we obtain

$$\frac{1}{|x[k]|^2}\boldsymbol{U}_L^H[k]\Big(|x[k]|^2\boldsymbol{R}_{\text{B}}[k]\Big)\boldsymbol{U}_L[k] = \boldsymbol{\Lambda}_L[k] \ , \tag{89}$$

$$\frac{1}{|x[k]|^2}\boldsymbol{U}_L^H[k]\Big(|x[k]|^2\boldsymbol{R}_{\text{D}}[k]\Big)\boldsymbol{U}_L[k] = \boldsymbol{I}_L \ . \tag{90}$$

In other words, the eigenvalue matrix $\boldsymbol{\Lambda}_L[k]$ does not change with respect to the input signal $x[k]$, even though the eigenvector matrix $\boldsymbol{U}_L[k]$ is scaled by $1/|x[k]|$. We can rewrite $\underline{\boldsymbol{u}}_v$ and $\underline{\boldsymbol{r}}_{\text{B}}$ in (71) as

$$\underline{\boldsymbol{u}}_v = \begin{bmatrix}\boldsymbol{0}_L^T & \cdots & \boldsymbol{u}_g'[k] & \cdots & \boldsymbol{0}_L^T\end{bmatrix}^T \ , \tag{91}$$

$$\underline{\boldsymbol{r}}_{\text{B}} = \begin{bmatrix}\boldsymbol{r}_{\text{B}}'^T[0] & \cdots & \boldsymbol{r}_{\text{B}}'^T[K-1]\end{bmatrix}^T \ , \tag{92}$$

where

$$\boldsymbol{u}_g'[k] = \frac{\boldsymbol{u}_g[k]}{|x[k]|} \ , \tag{93}$$

$$\boldsymbol{r}_{\text{B}}'[k] = |x[k]|^2\boldsymbol{r}_{\text{B}}[k] \ , \tag{94}$$

where $\boldsymbol{u}_g'[k]$ is the gth eigenvector at the kth block, then we obtain

$$\begin{aligned}\underline{\boldsymbol{u}}_v^H\underline{\boldsymbol{r}}_{\text{B}}\underline{\boldsymbol{u}}_v &= \boldsymbol{u}_g'^H[k]\boldsymbol{r}_{\text{B}}'[k]\boldsymbol{u}_g'[k] \\ &= \boldsymbol{u}_g^H[k]\boldsymbol{r}_{\text{B}}[k]\boldsymbol{u}_g[k] \ ,\end{aligned} \tag{95}$$

and λ_v is independent of the input signal. If V and μ_{bf} are selected by the user, then the broadband approach VAST-BF is also signal-independent as in VAST-NF. This result can be seen as one of the differences between the DFT-domain VAST and the time-domain VAST because the DFT-domain approach cannot directly consider the input signal.

Superficially, VAST-BF seems to be the same approach as VAST-NF, with the only obvious difference appearing to be whether it is one big problem or K small problems that must be solved. However, VAST-BF is clearly different from VAST-NF. To explain this more clearly, one example is depicted in Fig. 3. If we compute the joint diagonalization for the kth and k'th frequency bins, respectively, then we obtain $2L$ eigenvalues and $2L$ eigenvectors, which are locally sorted in descending order, such that the order of the eigenvalues is $(\lambda_1[k], \cdots, \lambda_L[k], \lambda_1[k'], \cdots, \lambda_L[k'])$. If VAST-NF, $1 \leq V \leq L$ eigenvalues for the kth and k'th frequency bins are respectively selected, i.e., $(\lambda_1[k], \cdots, \lambda_V[k], \lambda_1[k'], \cdots, \lambda_V[k'])$. In

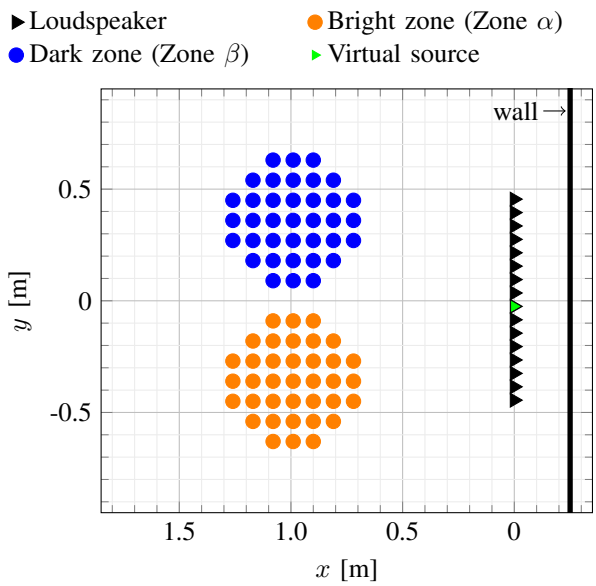


Fig. 4. The system geometry. The number of loudspeakers (black triangle) $L = 16$, the number of control points (orange and blue circles) $M = 37$ for each zone, and the virtual source (green triangle) are shown.

contrast, these $2L$ eigenvalues are sorted in a globally descending order for VAST-BF, such that the first $2 \leq 2V \leq 2L$ eigenvalues are chosen; therefore, this is not always the same order as that of VAST-NF. If all L eigenvalues in the kth frequency bin are smaller than those in the k'th frequency bin, then nothing is selected from the kth frequency bin, which means that no control effort is assigned in the k'th frequency bin. For example, if $V \leq 3$ in Fig. 3, then only the eigenvalues from the second block are selected, and so are the eigenvectors. It should be noted that the control filter of VAST-BF with $V = 1$ only passes the one frequency that results in the maximum acoustic contrast and filters the rest out, which is the same behavior as BACC. The same behavior is observed if only one frequency bin is considered for VAST-NF.

V. Experimental Validation and Discussion

In this section, the proposed methods are validated by three scientific questions as follows:

- How does the trade-off between oAC and SDE behave for the proposed methods?
- How does the performance change with respect to the two user parameters V and μ?
- How does the performance vary between the different methods and different domains?

The validation process is based on the narrowband approach VAST-NF, although this process can also be applied for the broadband approach VAST-BF as well as the time-domain approach VAST-T without the loss of generality. Throughout this section, frequency bin index k is omitted unless otherwise specified. Note that while we here study a particular setup with two sound zones, this is done to provide a fundamental understanding of the proposed methods. Finally, all of the performance metrics are represented in dB scale[4].

[4]The MATLAB code is available at https://github.com/actlee/vast_dft

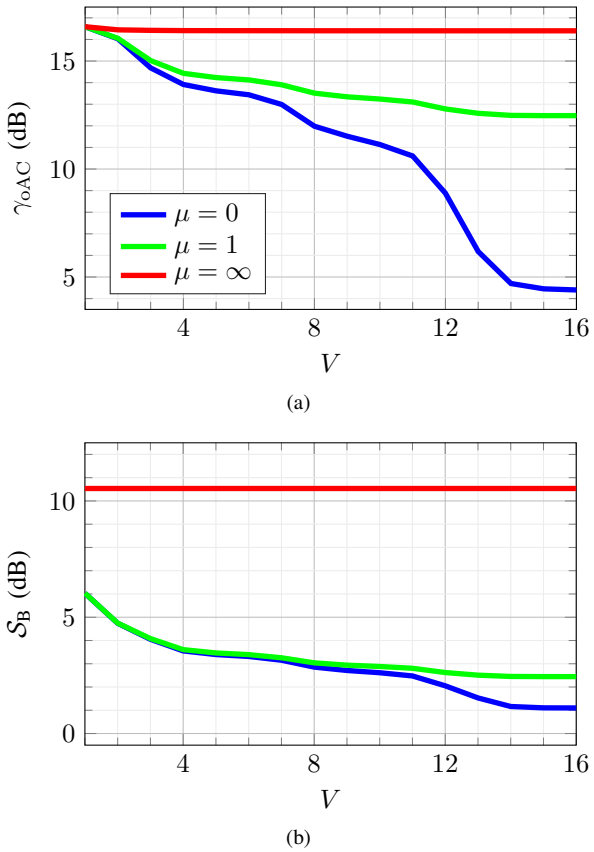


Fig. 5. (a) The acoustic contrast γ_{oAC} in dB and (b) the signal distortion energy $\mathcal{S}_{\mathrm{B}}$ in dB as a function of V for three different μ: $\mu = 0$ (blue), $\mu = 1$ (green), and $\mu = \infty$ (red).

A. System setup

As illustrated in Fig. 4, a uniform line array with $L = 16$ and a spacing between the adjacent loudspeakers of $d = 0.09$ m was considered. The zones were located in front of the loudspeaker array, and each of the zones was spatially sampled by $M = 37$ control points. The same distance as that between the loudspeakers was used for the control points. The speed of sound was taken as 343 m/s, the sampling frequency was set to $f_s = 16$ kHz, and the length of the control filters was set to $J = 240$ samples which corresponds to 15 ms at f_s and to a frequency grid of 66.67 Hz. The number of frequency bins was $K = 121$, including the DC and Nyquist frequency bins. Spatial aliasing occurs approximately at 1.9 kHz. A room with the dimensions $4.5 \times 4.5 \times 2.2$ m and a reverberation time $T_{60} = 300$ ms was considered. The measured room impulse responses (RIRs) $\hat{h}_{ml}[n]$ and the system geometry used in [30], [52] were used in all of the experiments throughout this section except for Sec. V-C. The impulse response of the desired sound field $\hat{h}_{mz}[n]$ was chosen from the RIR of the 8th loudspeaker after being truncated to contain only the direct path component. The location of the virtual source was chosen for convenience. The RIRs were resampled from 48 kHz to 16 kHz, and each of them was of length $2,967$. Zones α and β were considered to be the bright and dark zones, respectively.

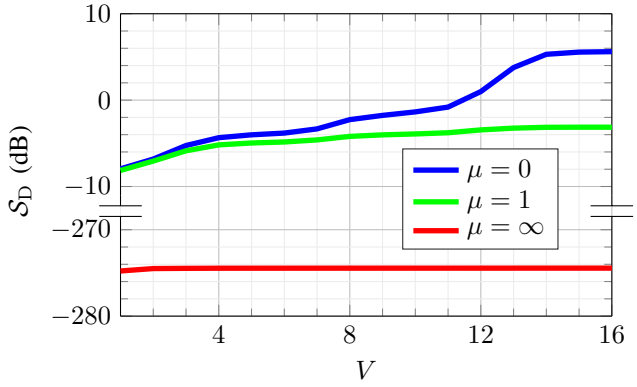


Fig. 6. The residual energy $\mathcal{S}_{\mathrm{D}}$ in dB as a function of V for three different μ: $\mu = 0$ (blue), $\mu = 1$ (green), and $\mu = \infty$ (red).

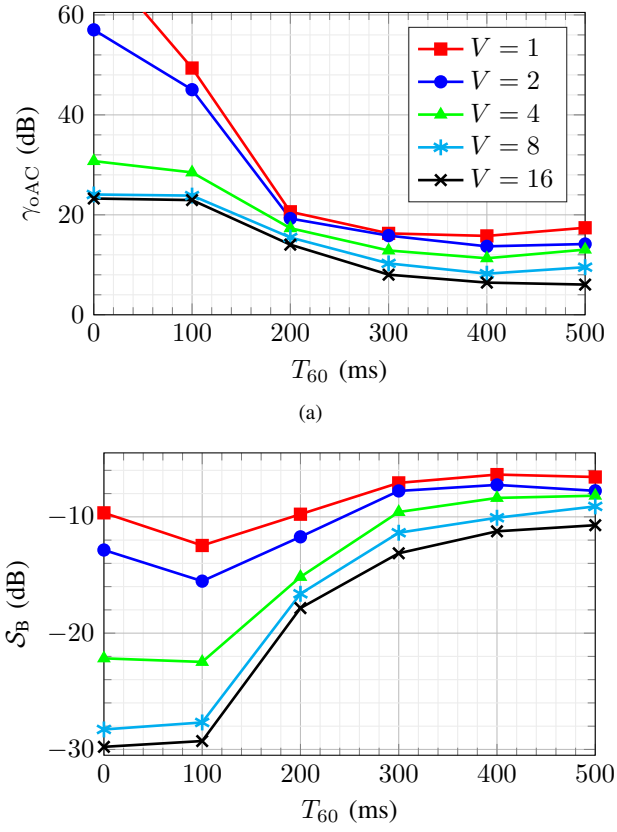
TABLE II
THE SUBSPACE RANK V AND THE OPTIMAL μ ACCORDING TO THE CONSTRAINT AND THE PERFORMANCE OF γ_{oAC}, η_{B}, AND η_{D}.

Subspace Rank V	8	16	8	16
Constraint	$\eta_{\mathrm{B}} \leq -7$ dB		$\eta_{\mathrm{B}} \leq -8$ dB	
Existence of μ^{opt}	YES	YES	NO	YES
μ^{opt}	3.2	4.2	0.0	1.1
oAC γ_{oAC} (dB)	14.7	12.0	14.8	12.7
nSDE η_{B} (dB)	−7.0	−7.0	−7.7	−8.0
nRE η_{D} (dB)	−35.4	−31.7	−35.7	−32.9

B. Trade-off between oAC and SDE

In this section, oAC γ_{oAC}, SDE $\mathcal{S}_{\mathrm{B}}$, and RE $\mathcal{S}_{\mathrm{D}}$ for VAST-NF are investigated. Note that the magnitude of the input signal is assumed to be unity, and the target frequency of 1 kHz is considered. First, oAC and SDE are shown in Fig. 5. When there is no control, the prior acoustic contrast (iAC) is $\gamma_{\mathrm{iAC}} = -0.5$ dB. For $V = 1$, the largest oAC $\gamma_{\mathrm{oAC}} = 16.6$ dB is always obtained, i.e., λ_1. On the other hand, the least oAC $\gamma_{\mathrm{oAC}} = 4.5$ dB is obtained when $V = 16$ and $\mu = 0$, but still about a 5 dB improvement on the acoustic contrast from iAC is observed. A similar trend as oAC occurs for SDE, as shown in Fig. 5 (b). When $\mu \to \infty$, the largest SDE is returned despite V. Therefore, we can clearly observe a trade-off between oAC and SDE by adjusting V and μ, which leads us to conclude that the largest oAC and the smallest SDE cannot be obtained at the same time. The cases of $\mu = 0$ and $\mu \to \infty$, respectively, provide the lower and the upper bounds as functions of V for oAC and SDE, as explained in (80) and (76), also respectively.

This trade-off influences RE $\mathcal{S}_{\mathrm{D}}$. The upper and lower bounds, respectively, are obtained for $\mu = 0$ and $\mu \to \infty$, which have the opposite trend of oAC and SDE. Approximately -275 dB (≈ 0) of RE can be achieved for $\mu \to \infty$; however, this is only possible for cases in which the magnitude of the control filter is almost equal to 0, as alluded to in Sec. IV-B. In other words, if $\mu \to \infty$, then minimizing RE is the only cost that the optimization problem tries to achieve; as a result, the performance of oAC and SDE is maximized, whereas RE is minimized, as seen in Figs. 5 and 6.



(a)

(b)

Fig. 7. (a) The acoustic contrast γ_{oAC} and (b) the signal distortion $\mathcal{S}_{\mathrm{B}}$ in dB as a function of the reverberation time T_{60} in ms for a fixed $\mu = 1$ and five different V: $V = 1$ (red), $V = 2$ (blue), $V = 4$ (green), $V = 8$ (cyan), and $V = 16$ (black).

C. Performance with respect to acoustic environment

In this section, we investigate the performance of oAC and SDE with respect to different acoustic environments. To simulate such environments, the RIRs corresponding to different T_{60} were calculated using the RIR generator toolbox [53], which is a MATLAB implementation of the image source method [54]. The same system setup, as described in the previous section, was used except for the RIRs. For the two user parameters, a fixed $\mu = 1$ and five different V were chosen; therefore, the oAC and SDE scores of ACC [11] and PM [13] can also be observed for $V = 1$ and $V = 16$, respectively. As can be seen in Fig. 7, oAC at each V decreases for an increasing T_{60}, whereas SDE increases for an increasing T_{60}. In Fig. 7 (a), the highest and lowest oAC at each T_{60} are always obtained for $V = 1$ and $V = 16$, respectively. A similar trend is observed for SDE, as shown in Fig. 7 (b). In other words, the largest and smallest SDE at each T_{60} are also observed for $V = 1$ and $V = 16$, respectively. If the given acoustic environment is more reverberant, then the performance of oAC and SDE degrades compared to the anechoic environment, which means that the control is more challenging due to the reverberations.

D. Performance with respect to the two user parameters

In Sec. V-B, the upper and lower bounds as functions of V for oAC, SDE, and RE were observed. In this section, nSDE

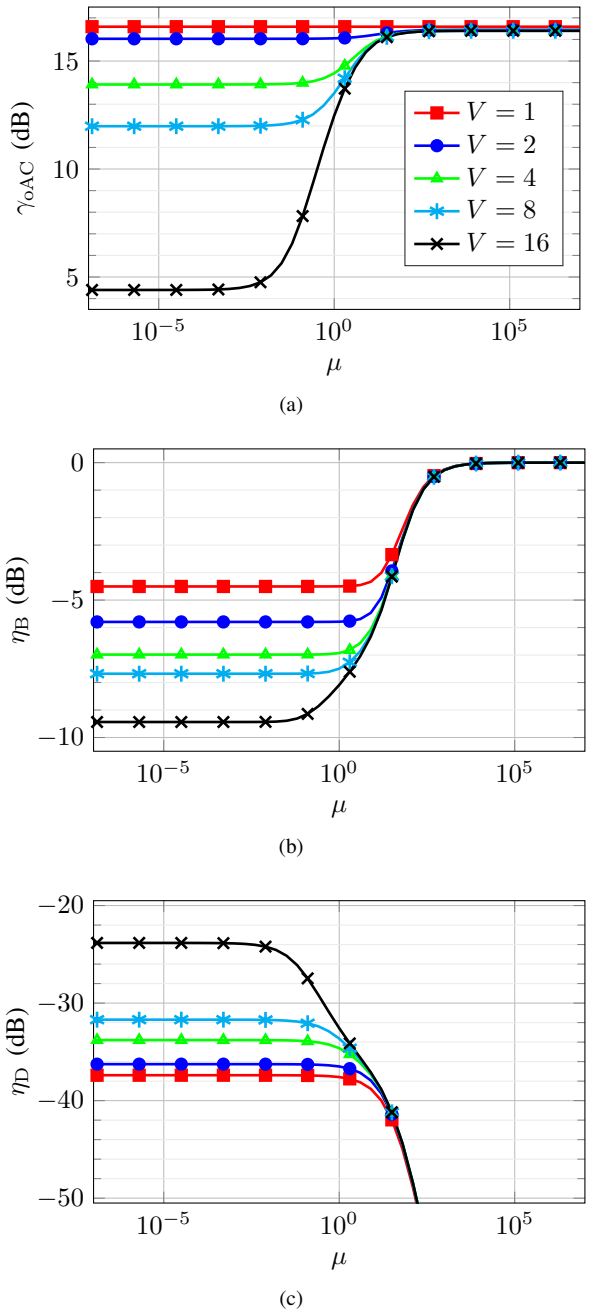


Fig. 8. (a) The acoustic contrast γ_{oAC} in dB, (b) the normalized signal distortion energy η_{B} in dB, and (c) the normalized residual energy η_{D} in dB as a function of μ for five different V: $V = 1$ (red), $V = 2$ (blue), $V = 4$ (green), $V = 8$ (cyan), and $V = 16$ (black).

and nRE are used instead of SDE and RE to investigate the signal-independent performance. In other words, the performance is only determined by the system geometry and the user parameters. The system setup with the RIRs described in Sec. V-A was used, and we will use this in the remainder of this section.

We observe the performance of oAC, nSDE, and nRE with respect to the two user parameters, as depicted in Fig. 8. Because all of the performance boundaries for different metrics are frequency dependent for VAST-NF, a frequency-dependent strategy for the constraint is necessary if one expects to achieve the optimal performance for the given system and the frequency of interest. However, this is not the case for VAST-BF and VAST-T because one global constraint is considered.

First, as V increases with a fixed μ, we can observe the trends in oAC, nSDE, and nRE. The upper and lower bounds as functions of μ for these metrics can be seen in Fig. 8. It should be noted that oAC and nSDE converge to the upper bound, while nRE converges to the lower bound regardless of V if μ is sufficiently large, e.g., $\mu > 10^3$. As alluded to in [39], μ is a user parameter that can be adjusted by the user; however, in this case the constraint might often be difficult to satisfy. If so, sub-optimal performance might result. One of the root finding methods, e.g., Newton's method, could be exploited as in [42] to find the optimal μ^{opt}, which would require the target ϵ as well as an initial value of μ_0 to be assigned; otherwise, finding the root cannot be done properly. Fortunately, these steps also can be carried out by using the graphs shown in Fig. 8, which guides the selection of μ visually, because there is a single constraint on the energy either in the bright or dark zone. For instance, if we consider (87) and place two different constraints respectively on nSDE, i.e., $\eta_{\mathrm{B}} \leq -7$ dB and $\eta_{\mathrm{B}} \leq -8$ dB, then μ is determined with dependence on V. Note that $\mu_0 = 0$ is a reasonable initial value because this guarantees the minimum nSDE. The optimal values μ^{opt} for $V = 8$ and $V = 16$ and corresponding performance are summarized in Table II. If $V \leq 8$ for $\eta_{\mathrm{B}} \leq -8$ dB, then none of μ can fulfill this constraint; hence, μ_0 should be considered as the optimal value μ^{opt} that gives the minimum nSDE for a given V.

E. Performance comparison between VAST-NF and VAST-BF

In this section, VAST-NF and VAST-BF are compared using the optimization criteria explained in Sec. IV-C. The subspace rank $V = L/4$ for K frequency bins is used for VAST-NF, whereas the subspace rank $V = LK/4$ is used for VAST-BF. Note that in total, both of the methods use the same number of eigenvectors and eigenvalues, i.e., $LK/4$. The values of oAC γ_{oAC}, nSDE η_{B}, and nRE η_{D} with respect to the frequency for different optimization problems are shown in the left, middle, and right columns in Fig. 9, respectively.

First, the optimization problem (85) in Sec. IV-C1 with the constraint $\mathcal{S}_{\mathrm{D}} \leq 13$ dB is considered for VAST-NF for all $K = 121$ frequency bins and VAST-BF for one global value. As shown in Figs. 9 (a) and (b), it is observed that, in general, VAST-BF has a higher oAC as well as a higher nSDE than that of VAST-NF in the frequency range less than 2 kHz. On the other hand, a lower nRE is observed in VAST-BF compared to VAST-NF in the same frequency range, as depicted in Fig. 9 (c).

Second, more precise control over nSDE can be obtained if the optimization problem (86) in Sec. IV-C2 is considered. The constraint $\eta_{\mathrm{D}} \leq -37$ dB is chosen for VAST-NF for all $K = 121$ frequency bins and VAST-BF for one global value. The same trend as in the previous experiment for both of the methods is observed from the values of oAC and nSDE, as shown in Figs. 9 (d) and (e), respectively. However, nRE performed by VAST-NF was kept $\eta_{\mathrm{D}} \leq -37$ dB for all K frequency bins as shown in Fig. 9 (f), whereas this constraint is applied on average across all K frequency bins in VAST-BF.

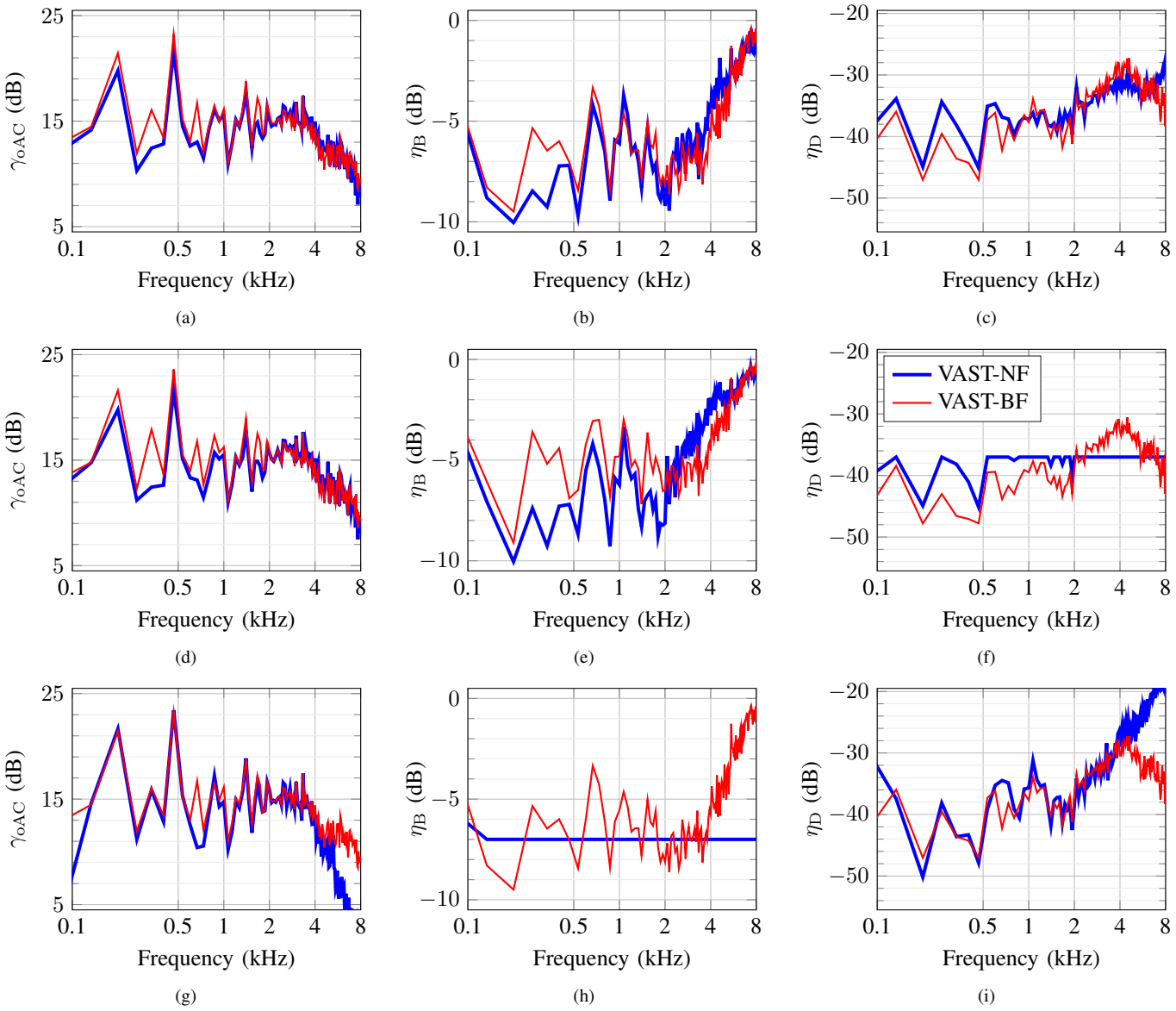


Fig. 9. The results of the optimization problems in Sec. IV-C. The left column shows the acoustic contrast γ_{oAC} in dB, the middle column shows the normalized signal distortion energy η_{B} in dB, and the right column shows the normalized residual energy η_{D} in dB as a function of the frequency bin for the proposed methods: VAST-NF (blue) with $(V, \mu) = (4, 1)$ for 121 frequency bins and VAST-BF (red) with $(V, \mu) = (484, 1)$, (a) to (c) the optimization problem (85) in Sec. IV-C1 and the constraint $\mathcal{S}_{\mathrm{D}} \leq 13$ dB, (d) to (f) the optimization problem (86) in Sec. IV-C2 and the constraint $\eta_{\mathrm{D}} \leq -37$ dB, and (g) to (i) the optimization problem (87) in Sec. IV-C3 and the constraint $\eta_{\mathrm{B}} \leq -7$ dB.

Lastly, the optimization problem (87) in Sec. IV-C3 with the constraint $\mathcal{S}_{\mathrm{B}} \leq -7$ dB is considered for VAST-NF for all $K = 121$ frequency bins and VAST-BF for one global value. Relatively similar performance of AC is observed from both of the methods in Figs. 9 (g), whereas the values of nSDE are precisely controlled in VAST-NF, as shown in Fig. 9 (h). A lower nRE is observed in VAST-BF compared to VAST-NF in Fig. 9 (h) in which a higher nSDE is observed, e.g., in the frequency range approximately more than 4 kHz.

As can be seen in these experiments, precise control across frequencies is possible using VAST-NF with a certain constraint at each frequency bin. If one has specific target constraints for different frequency bins, then VAST-NF would be preferable. It should be noted that exactly the same solution is obtained from these two methods when $V = LK$ and $\mu_{\mathrm{nf}} = \mu_{\mathrm{bf}}$. If $1 \leq V \leq K$ for VAST-BF, then some frequency bins cannot be controlled, depending on the magnitude of their eigenvalues. However, VAST-NF always assigns at least one eigenvector corresponding to the largest eigenvalue at each frequency bin; hence, all K frequency bins can be controlled. For VAST-BF, the subspace rank V at each frequency bin will be determined based on the magnitude of the eigenvalues across frequencies. At certain frequency bins in which a smaller V (positive integer) is determined, as seen in Fig. 10, a higher oAC by VAST-BF than that by VAST-NF would be generally obtained when the same constraint is considered for both of the methods. Specifically, in these experiments, it is observed in the frequency range less than 2 kHz, as shown in

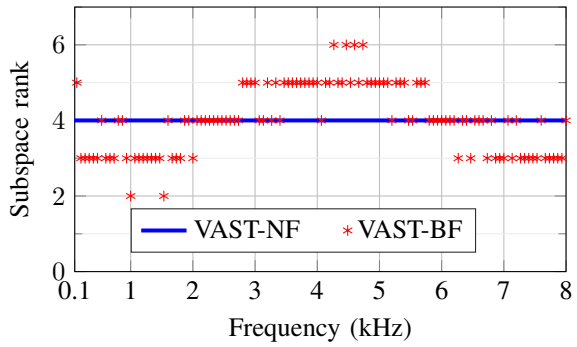


Fig. 10. The subspace rank as a function of frequency bin for VAST-NF (blue) and VAST-BF (red). The optimization problem considered for Figs. 9 (d) to (f) is used to get the results in this figure.

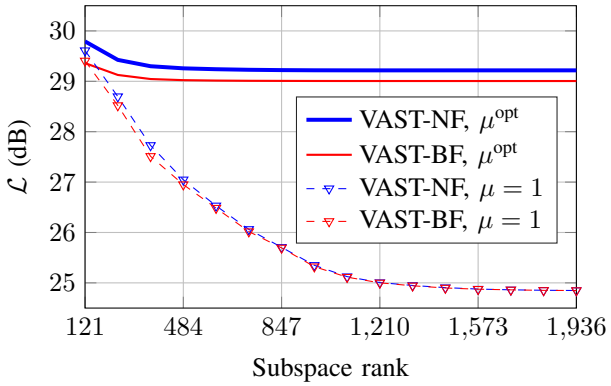


Fig. 11. The cost function $\mathcal{L}$ in dB as a function of subspace rank for the proposed methods: VAST-NF (blue) and VAST-BF (red).

the left column in Fig. 9. This observation leads us to conclude that VAST-BF can implicitly move oAC from one frequency to another frequency, which VAST-NF, on the other hand, cannot do. We remark that it was shown that the perceptual effect of this property can be represented as a weighting corresponding to a time-varying filter and incorporated into the VAST-T framework, leading to perceptually weighted reproduction errors and acoustic contrast, which allows us to reshape the reproduction errors accordingly [41]. Sorting the eigenvalues either globally or locally leads us to this significant difference. Note that this property depends on the spatial statistics determined by the geometry of the loudspeaker array and the locations of the sound zones.

Finally, the cost functions of VAST-NF and VAST-BF are compared. The same optimization problem with the same constraint used in Figs. 9 (d) to (f) is considered. Note that the comparison is only valid when at least $V = K$ for VAST-BF and $V = 1$ for K frequency bins for VAST-NF are used. For a fixed subspace rank, the cost functions for all K frequency bins are calculated, and their sum is plotted in Fig. 11. Because VAST-BF minimizes the cost function globally, it always gives the lower value of the cost function in comparison to that of VAST-NF. Even in the case of a fixed $\mu = 1$, the value of the cost function for VAST-BF is always lower than or equal to that of VAST-NF.

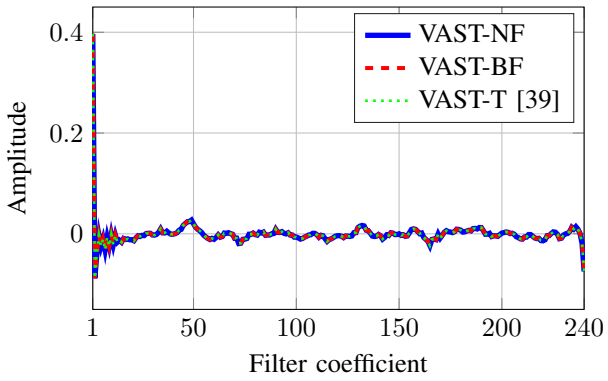


Fig. 12. The control filter of length $J = 240$ at the 8th loudspeaker using three different methods: VAST-NF (blue), VAST-BF (red dash), and VAST-T [39] (green dot).

TABLE III
THE MSE FOR DIFFERENT PAIRS OF THE CONTROL METHODS

Pair	VAST-NF VAST-BF	VAST-NF VAST-T	VAST-BF VAST-T
MSE	6.97e−30	6.16e−28	6.17e−28

F. Comparison between VAST-NF, VAST-BF, and VAST-T

Until now, we have investigated the two proposed methods in the frequency-domain. We extend this comparison by including the time-domain method, VAST-T. In some literature, e.g., [26], [43], the frequency- and time-domain approaches behave differently, whereas in other works, e.g., [44], they behave identically. In this section, we investigate this conflict. It should be noted that the solutions for VAST-NF and VAST-BF are calculated in the frequency-domain. The time-domain control filter of each method is then obtained with an inverse Fourier transform.

1) A special case: Fig. 12 shows the filter coefficients for the 8th loudspeaker for three different methods: VAST-NF, VAST-BF, and VAST-T [39]. The subspace rank V is chosen as the full rank for each of the methods, and μ is chosen as $\mu = 1$. The length of RIRs I is set to the same length as the control filter J and the input signal N, i.e., $I = J = N = 240$. Readily, we can conclude that VAST-NF and VAST-BF provide the same performance as described in Sec. V-E. The solutions from these three methods coincide with each other, which can be confirmed by the MSE for different pairs between the three methods summarized in Table III, under the following conditions: 1) $N = J \geq I$, and 2) the input signal $x[n]$ is periodic in N, as claimed in [39]. Note that in this case, the number of frequency bins K is equal to $N/2+1$. As long as these conditions hold true, the results from these methods are the same, although typically the frequency-domain approach is considered to be a faster version of the time-domain approach, e.g., [44]. VAST-NF is the most efficient method in terms of computational complexity in only this specific scenario that J is at least equal to or longer than I, because all three methods give exactly the same performance.

2) The general case: In the previous section, we compared the frequency- and time-domain approaches to show when

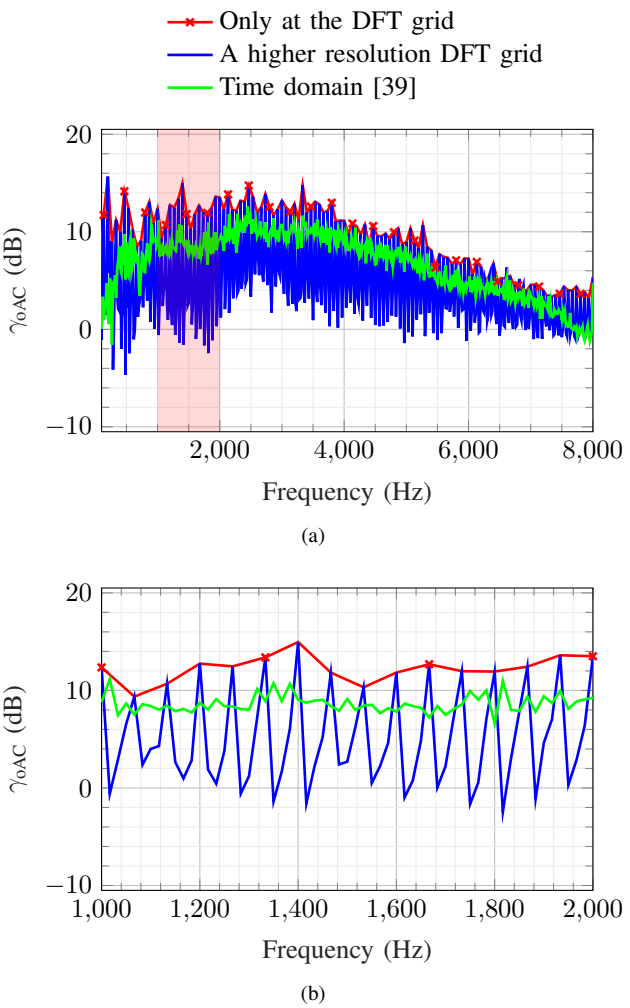


Fig. 13. (a) and (b) The acoustic contrast γ_{oAC} from the frequency-domain and the time-domain approaches: the frequency-domain approach only at the DFT grid (red), a higher resolution DFT grid (blue), and the time-domain approach, VAST-T [39] (green). Note that panel (b) is a magnified version of panel (a) for the frequency range from 1 kHz to 2 kHz.

those methods give the same performance. VAST-NF (and/or VAST-BF) is efficient and fast compared to VAST-T if all three methods give the same performance; however, this is only true under the two conditions described in the previous section. Therefore, the frequency-domain approach is an approximate and fast version of the time-domain approach because making the input signal N-periodic is not typically the case and is not typically possible. Because the input signal is N-periodic, the performance from the frequency-domain approaches is guaranteed at the DFT grid.

To validate this claim, we designed the following experiment with the same V and μ, as used in the previous section. The frequency-domain approach VAST-NF is considered in Fig. 13 because VAST-NF and VAST-BF behave identical in this case. VAST-T is shown as the time-domain approach in green in Fig. 13. Note that the Kronecker delta function is used as the input signal to calculate the control filter by VAST-T. A set of sinusoidal signals from 16.67 Hz to 8 kHz with a frequency grid of 16.67 Hz is considered to calculate oAC. In total, 480 sinusoidal signals are considered, and the length of each signal is set to 1 second.

The frequency-domain approach gives a fairly high oAC on the DFT grid, shown in red in Fig. 13. However, as soon as it is not on the DFT grid, the oAC decreases rapidly, as depicted in blue in Fig. 13. This phenomenon can also be observed in the literature, e.g., [20], [26], [27], [43]. This means that the frequency-domain approach can only control the performance of oAC on the DFT grid. In contrast, although the performance of the time-domain approach is lower than that of the frequency-domain approach on the DFT grid, the time-domain approach has a more stable performance across frequencies.

VI. Conclusion

In this paper, we proposed two methods for the creation of a bright zone and a dark zone: a narrowband approach (VAST-NF) and a broadband approach (VAST-BF). This process can be implemented in the frequency-domain by using variable span trade-off filters. To create sound zones, a set of loudspeakers is used, and each of the zones is spatially discretized into a set of control points. Spatial correlation matrices of the bright and dark zones are computed and then are jointly diagonalized by using a generalized eigenvalue decomposition. Exploiting the joint diagonalization to make the spatial correlation matrices low-rank, a flexible framework that can have a trade-off between acoustic contrast and signal distortion is obtained. This solution is obtained by solving a convex optimization with a constraint.

The two proposed methods are thoroughly investigated both theoretically and experimentally. VAST-BF shows that the acoustic contrast can be transferred from one frequency bin to another frequency bin depending on the subspace rank, which is not the case for VAST-NF. Furthermore, the two proposed methods are also compared to the time-domain approach (VAST-T) with respect to the computational complexity and the acoustic contrast. We show that VAST-T can be exactly the same as VAST-NF and VAST-BF only if the input signal is periodic and the control filter is equal to or longer than the room impulse response. In general, VAST-NF and VAST-BF are faster than VAST-T and not computationally complex and guarantee a fairly good performance on the DFT grid, whereas VAST-T provides a more stable acoustic contrast across frequencies but has the most expensive computational complexity.

Acknowledgment

The authors would like to thank Dr.-Ing. M. Schneider and Assoc. Prof. E. A. P. Habets for providing the RIRs.

References

[1] W. F. Druyvesteyn, R. M. Aarts, A. Asbury, P. Gelat, and A. Ruxton, "Personal sound," in *Proc. Inst. Acoust.*, vol. 16, no. 2, 1994, pp. 571–585.
[2] W. F. Druyvesteyn and J. Garas, "Personal sound," *J. Audio Eng. Soc.*, vol. 45, no. 9, pp. 685–701, Sep. 1997.
[3] T. Betlehem, W. Zhang, M. A. Poletti, and T. D. Abhayapala, "Personal sound zones: Delivering interface-free audio to multiple listeners," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 81–91, Mar. 2015.
[4] S. J. Elliott and M. Jones, "An active headrest for personal audio," *J. Acoust. Soc. Am.*, vol. 119, no. 5, pp. 2702–2709, May 2006.

[5] S. Widmark, "Causal MSE-optimal filters for personal audio subject to constrained contrast," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 27, no. 5, pp. 972–987, May 2019.

[6] H. So and J.-W. Choi, "Subband optimization and filtering technique for practical personal audio systems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Brighton, UK, May 2019, pp. 8494–8498.

[7] J.-H. Chang and W.-H. Cho, "Evaluation of independent sound zones in a car," in *23rd Int. Congr. Acoust.*, Aachen, Germany, Sep. 2019.

[8] J. Brunskog, F. M. Heuchel, D. C. Nozal, M. Song, F. T. Agerkvist, E. F. Grande, and E. Gallo, "Full-scale outdoor concert adaptive sound field control," in *23rd Int. Congr. Acoust.* Aachen, Germany: German Acoustical Society (DEGA), Sep. 2019, pp. 1170–1177.

[9] S. J. Elliott, H. Murfet, and K. R. Holland, "Minimally radiating arrays for mobile devices," in *Proc. 16th Int. Congr. Sound Vib.*, Krakow, Poland, Jul. 2009.

[10] J. Cheer, S. J. Elliott, Y. Kim, and J.-W. Choi, "Practical implementation of personal audio in a mobile device," *J. Audio Eng. Soc.*, vol. 61, no. 5, pp. 290–300, Jun. 2013.

[11] J.-W. Choi and Y.-H. Kim, "Generation of an acoustically bright zone with an illuminated region using multiple sources," *J. Acoust. Soc. Am.*, vol. 111, no. 4, pp. 1695–1700, 2002.

[12] O. Kirkeby and P. A. Nelson, "Reproduction of plane wave sound fields," *J. Acoust. Soc. Am.*, vol. 94, no. 5, pp. 2992–3000, Nov. 1993.

[13] M. A. Poletti, "An investigation of 2D multizone surround sound systems," in *Proc. 125th Conv. Audio Eng. Soc.*, San Francisco, CA, USA, Oct. 2008, Paper 7551.

[14] Y. J. Wu and T. D. Abhayapala, "Theory and design of soundfield reproduction using continuous loudspeaker concept," *IEEE Trans. Audio, Speech, Language Process.*, vol. 17, no. 1, pp. 107–116, Jan. 2009.

[15] ——, "Spatial multizone soundfield reproduction: Theory and design," *IEEE Trans. Audio, Speech, Language Process.*, vol. 19, no. 6, pp. 1711–1720, Aug. 2011.

[16] W. Zhang, T. D. Abhayapala, T. Betlehem, and F. M. Fazi, "Analysis and control of multi-zone sound field reproduction using modal-domain approach," *J. Acoust. Soc. Am.*, vol. 140, no. 3, pp. 2134–2144, Sep. 2016.

[17] P. Coleman, P. J. B. Jackson, M. Olik, and J. A. Pedersen, "Personal audio with a planar bright zone," *J. Acoust. Soc. Am.*, vol. 136, no. 4, pp. 1725–1735, Oct. 2014.

[18] J.-H. Chang and F. Jacobsen, "Sound field control with a circular double-layer array of loudspeakers," *J. Acoust. Soc. Am.*, vol. 131, no. 6, pp. 4518–4525, Jun. 2012.

[19] J.-W. Choi, "Sound sketch: Shaping sound in space and time using loudspeaker arrays," in *Proc. Inter-Noise*, Melbourne, Australia, Nov. 2014.

[20] A. H. Andersen, "On multiple sound zones for wideband signals," Master's thesis, Aalborg University, Jan. 2014.

[21] Y. Cai, M. Wu, and J. Yang, "Sound reproduction in personal audio systems using the least-squares approach with acoustic contrast control constraint," *J. Acoust. Soc. Am.*, vol. 135, no. 2, pp. 734–741, Feb. 2014.

[22] M. Shin, S. Q. Lee, F. M. Fazi, P. A. Nelson, D. Kim, S. Wang, K. Park, and J. Seo, "Maximization of acoustic energy difference between two spaces," *J. Acoust. Soc. Am.*, vol. 128, no. 1, pp. 121–131, Jul. 2010.

[23] S. J. Elliott, J. Cheer, J.-W. Choi, and Y. Kim, "Robustness and regularization of personal audio systems," *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 7, pp. 2123–2133, Sep. 2012.

[24] Q. Zhu, P. Coleman, X. Qiu, M. Wu, J. Yang, and I. Burnett, "Robust personal audio geometry optimization in the svd-based modal domain," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 27, no. 3, pp. 610–620, Mar. 2019.

[25] S. J. Elliott and J. Cheer, "Regularisation and robustness of personal audio systems," ISVR Technical Memorandum 995, Tech. Rep., 2011.

[26] Y. Cai, M. Wu, and J. Yang, "Design of a time-domain acoustic contrast control for broadband input signals in personal audio systems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Vancouver, BC, Canada, May 2013, pp. 341–345.

[27] Y. Cai, M. Wu, L. Liu, and J. Yang, "Time-domain acoustic contrast control design with response differential constraint in personal audio systems," *J. Acoust. Soc. Am.*, vol. 135, no. 6, pp. EL252–EL257, Jun. 2014.

[28] D. H. M. Schellekens, M. B. Møller, and M. Olsen, "Time domain acoustic contrast control implementation of sound zones for low-frequency input signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Shanghai, China, Mar. 2016, pp. 365–369.

[29] M. Buerger, C. Hofmann, C. Frankenbach, and W. Kellermann, "Multi-zone sound reproduction in reverberant environments using an iterative least-squares filter design method with a spatiotemporal weighting function," in *Proc. IEEE Workshop Appl. of Signal Process. to Aud. and Acoust.*, New Paltz, NY, USA, Oct. 2017.

[30] M. Schneider and E. A. P. Habets, "Iterative DFT-domain inverse filter optimization using a weighted least-squares criterion," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 27, no. 12, pp. 1957–1969, Dec. 2019.

[31] M. B. Møller, M. Olsen, and F. Jacobsen, "A hybrid method combining synthesis of a sound field and control of acoustic contrast," in *Proc. 132nd Conv. Audio Eng. Soc.*, Budapest, Hungary, Apr. 2012, Paper 8627.

[32] M. F. Simón Gálvez, S. J. Elliott, and J. Cheer, "Time domain optimization of filters used in a loudspeaker array for personal audio," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 23, no. 11, pp. 1869–1878, Nov. 2015.

[33] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," *IEEE Transactions on Signal Processing*, vol. 50, no. 9, pp. 2230–2244, Sep. 2002.

[34] Y. Ephraim and H. L. Van Trees, "A signal subspace approach for speech enhancement," *IEEE Trans. Speech Audio Process.*, vol. 3, no. 4, pp. 251–266, Jul. 1995.

[35] S. H. Jensen, P. C. Hansen, S. D. Hansen, and J. A. Sørensen, "Reduction of broad-band noise in speech by truncated QSVD," *IEEE Trans. Speech Audio Process.*, vol. 3, no. 6, pp. 439–448, Nov. 1995.

[36] R. Serizel, M. Moonen, B. van Dijk, and J. Wouters, "Low-rank approximation based multichannel Wiener filter algorithms for noise reduction with application in cochlear implants," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 22, no. 4, pp. 785–799, Apr. 2014.

[37] J. R. Jensen, J. Benesty, and M. G. Christensen, "Noise reduction with optimal variable span linear filters," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 24, no. 4, pp. 631–644, Apr. 2016.

[38] J. K. Nielsen, T. Lee, J. R. Jensen, and M. G. Christensen, "Sound zones as an optimal filtering problem," in *Proc. 52th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Oct. 2018, pp. 1075–1079.

[39] T. Lee, J. K. Nielsen, J. R. Jensen, and M. G. Christensen, "A unified approach to generating sound zones using variable span linear filters," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Calgary, AB, Canada, Apr. 2018, pp. 491–495.

[40] T. Lee, J. K. Nielsen, and M. G. Christensen, "Towards perceptually optimized sound zones: A proof-of-concept study," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2019, pp. 136–140.

[41] ——, "Signal-adaptive and perceptually optimized sound zones with variable span trade-off filters," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 28, pp. 2412–2426, 2020.

[42] L. Shi, T. Lee, L. Zhang, J. K. Nielsen, and M. G. Christensen, "A fast reduced-rank sound zone control algorithm using the conjugate gradient method," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2020, pp. 436–440.

[43] M. B. Møller and M. Olsen, "Sound zones: On performance prediction of contrast control methods," in *Proc. Audio Eng. Soc. Int. Conf. Sound Field Control*, Guildford, UK, Jul. 2016.

[44] L. Vindrola, M. Melon, J.-C. Chamard, B. Gazengel, and G. Plantier, "Personal sound zones: A comparison between frequency and time domain formulations in a transportation context," in *Proc. 147th Conv. Audio Eng. Soc.*, New York, NY, USA, Oct. 2019, Paper 10216.

[45] L. Shi, T. Lee, J. K. Nielsen, and M. G. Christensen, "Subspace-based methods for the generation of personal sound zones with physically meaningful constraints," *IEEE/ACM Trans. Audio, Speech, Language Process.*, 2020, under review.

[46] P. A. Nelson, F. Orduna-Bustamante, and H. Hamada, "Inverse filter design and equalization zones in multichannel sound reproduction," *IEEE Trans. Speech Audio Process.*, vol. 3, no. 3, pp. 185–192, May 1995.

[47] O. Kirkeby and P. A. Nelson, "Digital filter design for inversion problems in sound reproduction," *J. Audio Eng. Soc.*, vol. 47, no. 7/8, pp. 583–595, Jul. 1999.

[48] P.-A. Gauthier and A. Berry, "Adaptive wave field synthesis for sound field reproduction: Theory, experiments, and future perspectives," *J. Audio Eng. Soc*, vol. 55, no. 12, pp. 1107–1124, Dec. 2007.

[49] J. Benesty, M. G. Christensen, and J. R. Jensen, *Signal enhancement with Variable Span Linear Filters*, 1st ed. Springer, Feb. 2016, vol. 7.

[50] G. H. Golub and C. F. Van Loan, *Matrix computations*, 4th ed. The Johns Hopkins University Press, 2012.

[51] Q. Zhu, P. Coleman, M. Wu, and J. Yang, "Robust acoustic contrast control with reduced in-situ measurement by acoustic modeling," *J. Audio Eng. Soc.*, vol. 65, no. 6, pp. 460–473, Jun. 2017.

[52] M. Schneider and E. A. P. Habets, "An iterative least-squares design method for filters with constrained magnitude response in sound reproduction," in *Proc. 43rd German Annu. Conf. Acoust. — DAGA*, Kiel, Germany, Mar. 2017, pp. 1347–1350.
[53] E. A. P. Habets, "Room impulse response generator," Technische Universiteit Eindhoven, Tech. Rep., Sep. 2010, Ver. 2.1.20141124.
[54] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.